Health Effects Division

Standard Evaluation Procedure

Reproductive Toxicity Studies

Health Effects Division
Office of Pesticide Programs
United States Environmental Protection Agency

Draft
April, 1993

Prepared by the following:

Gary Burin, Ph.D., DABT
Laurence Chitlik, M.S., DABT
Susan L. Makris, M.S.
James N. Rowe, Ph.D.

The significant contributions of Dr. Quang Bui to early work on this document are gratefully acknowledged by the authors, as well as extensive comments received from Dr. Eric Clegg, Mr. Roger Gardner and Dr. David Anderson.

STANDARD EVALUATION PROCEDURE

PREAMBLE

This Standard Evaluation Procedure (SEP) is one of a set of new or updated guidance documents which explain the current operating procedures used to evaluate toxicological data submitted to the Office of Pesticide Programs. Based upon our previous experiences with the first set of SEPs, they are designed to ensure a fully updated and consistent treatment of major scientific topics and to provide interpretive policy guidance where appropriate. The Standard Evaluation Procedures have been coordinated with Agency Risk Assessment Guidelines (where appropriate) and will be used in conjunction with the revised Pesticide Assessment Guidelines. While the documents were developed to interpret the principles of scientific evaluation within the Office of Pesticide Programs, they are intended to be of a broad enough scientific base to be of use for the evaluation of all related studies and scientific data within the Agency. The Standard Evaluation Procedures serve as a valuable internal reference documents for OPP risk assessors and risk managers and will inform the public and regulated community of important considerations in the evaluation of test data for chemicals. These SEPs will improve upon the state-of-science within EPA, and, in conjunction with the revised Pesticide Assessment Guidelines, will lead to a more balanced and effective use of public and private resources applied to the regulation of pesticide chemicals in commerce.

Penelope Fenner-Crisp, Director
Health Effects Division

TABLE OF CONTENTS

TABLES

FIGURES

PREFACE

The purpose of this SEP is to provide guidance on:

- how to evaluate reproductive toxicity data submitted to the Office of Pesticide Programs (OPP) to support registration or re-registration of pesticides and
- how to use that evaluation to support assessment of potential risks from human exposures.

In addition to addressing these purposes, the SEP contains reference material and tables of parameters that are essential knowledge in evaluation of reproductive effects.

This SEP should be used in conjunction with the Pesticide Assessment Guidelines for reproductive toxicity (Subdivision F, Hazard Evaluation: Human and Domestic Animals, §83-4) (USEPA, 1982) and with guidance from senior scientists experienced in the review of reproductive toxicity studies. This SEP is not a rigid set of rules to follow in the evaluation of reproduction toxicity studies.

Criteria for data acceptability are described in Section II. Section III describes various aspects of study conducts, while Section IV discusses scientific principles used in evaluation of parental, reproductive and developmental toxicity end points as a basis for sound interpretations of reproductive toxicity study results. Section V provides guidance on scientific issues related to risk assessment and risk characterization. Section VI discusses the format and content of reproductive toxicity study reviews including the Data Evaluation Record (DER) and the Peer Review document.

The following are some of the most commonly discussed issues in reproductive toxicity reviews and the relevant sections of this SEP to read:

1. The relationship between parental (maternal, paternal) toxicity and reproductive toxicity; determination of NOELs for these effects--See Sections I.D. and V.B.1.a.
2. The relevance of developmental (teratology) and other types of studies to the reproductive toxicity study-- See Section I.E.
3. Data used for determining the exposure dosage (premating, gestation, lactation, all) and how should it be presented in the review (i.e., ppm vs mg/kg/day)-- See Section IV.A.4.
4. Definitions for the reproductive indices and their limitations--See Sections IV.A.10.a-j and IV.B.1-2.
5. Format for data be presentation in a reproductive study--See DER format in Section VI.A.
6. Value and use of historical control data--See Section IV.C.4.
7. Extent of data required to perform a reproductive risk assessment and format for presentation--See Section I.D.
8. Use of other reproductive studies which are not full multigenerational studies--See Section II.C.
9. The acceptability of reproductive studies received from other countries (OECD, Japan)--See Section II.C..
10. Criteria for rejecting a study because of an insufficient number of litters--See Section III.B.3.

I. INTRODUCTION

A. Background Information

The multigeneration reproduction study determines potential adverse effects of exposure to xenobiotics on the male and female reproductive systems, the conceptus, and to the neonate. It should be regarded as a bioassay for a wide variety of endpoints related to reproduction. The multigeneration reproduction study which is properly designed, conducted, and interpreted can detect effects on libido, germ cells, gametogenesis, fertilization, implantation, embryonic, fetal, neonatal growth and development, lactation, and postweaning growth and maturity. The multigeneration reproduction study also provides information regarding the direct toxicity of a chemical to the pregnant animal. Due to the profound physiological changes which occur during pregnancy, these effects may be much different than those observed in chronic and subchronic studies.

The Agency guidelines for reproductive toxicity risk assessment are intended to ensure a consistent approach to evaluations. Interpretation of endpoints related to reproductive toxicity should be viewed in the context of these guidelines. The most relevant guidelines on this topic are the (1) Guidelines for Male Reproductive Risk Assessment (USEPA, 1988a) the (2) Guidelines for the Health Assessment of Suspect Female Reproductive Toxicants USEPA, 1988b), and the (3) Guidelines for Reproductive Toxicity Risk Assessment (USEPA, 1993a draft) which will combine the 1988 Guidelines. These Agency risk assessment guidelines will provide the framework for the evaluation of the multigeneration reproduction study.

Special studies such as the continuous breeding protocol and studies routinely required in the Office of Pesticide Programs (OPP) such as the subchronic and chronic feeding studies, and developmental toxicity studies provide additional information which may be relevant to reproductive toxicity. Although the focus of this Standard Evaluation Procedure (SEP) is on the multigeneration reproduction study (primarily the two-generation study), aspects of other studies relevant to reproductive toxicity will be discussed here as well. In addition, because reproductive toxicants may not cause effects solely by affecting the integrity of reproductive tissues but may also induce genetic abnormalities, adverse developmental effects and other toxic effects, the reviewer is referred to other relevant guidelines such as the Guidelines for Mutagenicity Risk Assessment (USEPA, 1986b), the Guidelines for Carcinogen Risk Assessment (USEPA, 1986a), and the Guidelines for Developmental Toxicity Risk Assessment (USEPA, 1991c).

B. Definitions

For the purpose of this Standard Evaluation Procedure (SEP), the following definitions are noted. Some of these terms are presented for explanatory purposes only as they are not often used in the routine evaluation of the multigeneration reproduction study.

1. Reproductive toxicity - The occurrence of adverse effects specifically related to the reproductive system that may result from exposure to environmental agents. Reproductive toxicity may be expressed as direct chemical-induced alterations to the female or male reproductive organs or the related endocrine system, or as a secondary effect of general adverse effects of a chemical. The manifestation of such toxicity may include but not be limited to, adverse effects on onset of puberty, gamete production and transport, cyclicity, sexual behavior, fertility, gestation, parturition, lactation, or pregnancy outcomes; premature reproductive senescence; or modifications in other functions that are dependent on the integrity of the reproductive system. Reproductive toxicity may be identified in relation to the affected sex, i.e., male reproductive toxicity or female reproductive toxicity.

2. Systemic toxicity - The term systemic toxicity encompasses reproductive effects but is used here as it relates to the generalized effects of a toxicant on the entire physiology of the body (e.g., reductions in body weights, various clinical signs of toxicity, reduction in food consumption).

3. Female reproductive cycle - The periodic recurrence of events in the neuroendocrine and generative systems (hypothalamus, pituitary, uterus, ovaries, and accessory sexual structures) associated with estrus in lower mammals and menstruation in humans and nonhuman primates.

4. Estrous cycle - One of the two types of reproductive cycles; the menstrual cycles in primates, and cycles of sexual receptivity in other female mammals. The cycles are divided on the basis of ovarian activity into estrus (period of sexual activity and proximity to follicle rupture or ovulation), metestrus (early development of the corpus luteum), diestrus (mature corpus luteum), and proestrus (period of follicular activity).

5. Fertility - Ability to conceive and to produce offspring within a defined period of time. For litter-bearing species, fecundity (the number of offspring) is also a measure of fertility.

   a. Fertile - Having a level of fertility that is within or exceeds the normal range for that species.

   b. Subfertile - Having a level of fertility that is below the normal range for that species but not infertile.

c. Infertile - Lacking fertility for a specified period of time. The infertile condition may be temporary; permanent infertility is termed sterility.

6. Male reproductive system - Those processes and organs in the male that are involved directly in sexual behavior and procreation. For this document, these include the testes, epididymides, vas deferens, accessory sex glands, penis, pituitary, and hypothalamus. The anterior pituitary (adenohypophysis) contains the various gonadotropins (LH, FSH) and lactotropins (prolactin, somatotropin, chorionic somatomammotropins) responsible for regulation of the reproductive organs and tissues including mammary tissue (Williams, 1974).

C. When Required

The Office of Pesticide Programs requires a multigeneration study when the use of the pesticide may result in dietary exposure to the pesticide; its active ingredient(s), metabolite(s), or degradation product(s) (i.e., when tolerances or exemptions from tolerances are considered) and, for nonfood uses, if exposure is expected over a portion of the human lifespan "which is significant in terms of the frequency of exposure, magnitude of exposure, or the duration of exposure" or reproductive concerns arise from other studies. The establishment of a temporary tolerance for residues in food may require a multigeneration study, depending upon factors such as extent and duration of exposure, structure-activity concerns, and the results of subchronic, developmental, and mutagenicity studies. Part 158 (40 CFR) (Data Requirements for Registration, Part 158.340, Toxicology Data Requirements, Table (a), Footnote 14) (USEPA, 1992) states that at least an interim report from the first generation of multigeneration reproduction study is needed if the theoretical maximum residue contribution is greater than 50% of the maximum permitted intake.

Structural similarities to known reproductive toxicants may lead to the need for reproductive toxicity testing when it otherwise would not be required for nonfood uses. However, small differences in structure may lead to major differences in the potential for reproductive toxicity. Screening assays for reproductive toxicity, such as the the Chernoff-Kavlock screen (1982) or the reproduction screen included in the OECD Screening Information Data Sets (SIDS) testing battery (1992), are generally not requested in lieu of a multigeneration study but may trigger the need for further testing (See Section II.B.5). As noted elsewhere in this SEP, limited information regarding reproductive toxicity is available from mutagenicity, subchronic, chronic and developmental toxicity studies, and this information may also be considered in making the determination

of whether to require a multigeneration reproduction study.

D. Utility of the Reproduction Study in a Regulatory Setting -- Systemic versus Reproductive NOELs

Data obtained from the two-generation reproduction study provide information concerning the potential systemic and reproductive effects of agents resulting from in utero through adult exposure over two generations. Similar to other chronic studies, reproductive toxicity studies may provide endpoints of toxicological concern upon which the RfD may be established. In addition, reproductive findings observed in these studies may be utilized as triggers for special review. Under special circumstances, findings observed in the multigeneration study may also lead to special testing in order to more clearly define the toxicity observed.

The reproduction study examines a wide variety of reproductive endpoints which may be of potential concern. However, due to the limitations of study design, it is often not readily apparent as to whether effects are male- or female-mediated. Further investigations or modifications of the routine protocol would be required for such determinations. Thus, when an effect on reproduction is found in the multigeneration reproduction study, followup studies may be necessary to elucidate the effect and to examine, with more specific methodology, the endpoint that appears to be affected. As well, since it is Agency policy to regulate pesticides based upon reproductive or developmental effects equally for both sexes and not exclude workers of either sex from the work place due to sex related toxicity (a gender neutral policy), it is still necessary to identify the gender in which the effects are observed.

It is recognized that certain chemicals that cause marked systemic toxicity in animals can affect reproductive performance. Large reductions in food consumption, body weight gain, and food efficiency in the parental animal can concommitently have associated effects upon pup weight and/or litter size, lactation indices, and other end-points evaluated in a reproduction study. Therefore, it may be incorrect to assume that the pesticide being tested is a "selective" reproductive toxicant since the manifestations observed may be associated with systemic toxic effects. However, a NOEL based upon either systemic effects or reproductive effects can be used as the basis for regulation of the chemical. That is, the study may be considered the critical study during the RfD assessment, independent of whether the NOEL is based upon reproductive or systemic effects.

Therefore, although the primary intent of the study is to assess potential reproductive effects, it is often impossible to separate these effects from other manifestations of general/systemic toxicity within the study. Since the separation of effects into "primary reproductive effects" or as effects "potentially secondary to other toxic manifestations" is often difficult, the establishment of a NOEL for reproductive effects should be undertaken only when it is clear that there is selective reproductive toxicity. The decision to set a reproductive NOEL may be determined in either of two ways. The first would be by identifying a lower NOEL for alterations in reproductive tissues or parameters related to reproductive ability than for other manifestations of toxicity (systemic effects). The second method in which this could be determined would be with special studies which identify a mechanism of action of a compound that may interfere with reproductive functions (i.e., cell division in gametogenesis, disruption of endocrine system balances that regulate estrous or menstrual cycles, neurological effects that may alter mating behavior, see Section V.3). For example, chemicals which interfere with microtubule formation, such as benomyl and its carbendizim metabolite, are clearly toxic to rapidly dividing cells such as sperm (Gray et al., 1990). Although effects are seen at other sites as well, a preferential sensitivity to the male reproductive system can be explained based upon knowledge of the mechanism of toxicity. In these cases, a NOEL can be established for reproductive toxicity. However, when there is a reduction in body weight gain, food consumption, food efficiency or other nonreproductive toxicity associated with reduced pup viability or lactation indexes, and the weight and food consumption effects have NOELs less than or equal to those for reproductive effects, a reproductive toxicity NOEL need not be clearly established.

In a situation where the effect appears to be selective to the reproductive system, other available toxicity studies in the data base should be examined before finalizing this assessment. Chronic, subchronic and mutagenicity studies should be carefully reviewed in order to identify the target tissue and potential toxicity of the pesticide (see Section I.E.). Only after this assessment is completed can the reviewer make any determination as to the necessity for a separate reproductive toxicity NOEL.

E. Correlation with Other Relevant Data

Even when the determination of a separate reproduction NOEL is not an issue of concern, the multigeneration reproduction study should be evaluated in the context of all other relevant information including developmental toxicity studies, subchronic and chronic studies, mutagenicity studies, and metabolism and

pharmacokinetic data. These studies are often conducted in the same species and strain (the Sprague-Dawley rat) as is the multigeneration study, occasionally at similar dose levels. Organ weights and histopathology of the ovary, uterus, testes, epididymides, seminal vesicles, prostate gland, pituitary and hypothalamus in chronic and subchronic studies may provide indications of target organs associated with reproductive toxicity. This information should be carefully considered in conjunction with the multigeneration study findings. Data regarding other endpoints (e.g., organ weight and histopathology from nonreproductive organs, clinical observations, and body weight data) are useful in understanding the general toxicity of the test compound. As noted above, comparison of dose levels inducing reproductive and nonreproductive toxicity and an understanding of the association between the various manifestations of toxicity may allow a determination of whether a chemical is a selective reproductive toxicant rather than a chemical which has reproductive effects secondary to other toxic manifestations.

Testicular histopathology data from the subchronic and chronic studies should be compared carefully to the results of the multigeneration study (see de Kretser and Kerr, 1988, for a discussion of testicular histology and function). A careful description of testicular histology facilitates study interpretation and is necessary for the separation of artifacts from compound-induced effects. Cell staging or morphometric measurements also facilitate the interpretation of testicular histology (see Section IV.C.2.a).

Formalin fixation, combined with paraffin embedding of the testis may result in artifacts such as shrinkage, vacuoles, and clumping of nuclear material which can mask effects and impair meaningful interpretation. The lack of any reported effect on the testicular histology may simply reflect poor specimen preparation. Cell staging or morphometric measurements also facilitate the interpretation of testicular histology (Russell et al., 1990). Even when histopathological changes in the testes are observed in the subchronic or chronic studies, effects on fertility may or may not be observed at similar dose levels in the multigeneration study. The lack of sensitivity of the multigeneration study for the detection of effects on fertility limits the ability of the study to confirm reproductive toxicity indicated by histological changes in reproductive tissues.

Reproductive data (single or multigeneration) can be extremely useful in confirming findings of developmental toxicity from a given chemical or sometimes in identifying developmental effects which may not otherwise be observed after dosing during the period of major organogenesis (gestation days 6-15 in rats) as

is done when evaluating developmental toxicity. Continued exposure during the entire gestational period, which occurs in a reproductive toxicity study, potentially allows the entire period of fetal development to be affected. When developmental effects are noted in reproductive studies, it is reasonable to evaluate the necessity of performing developmental studies with extended exposure during gestation.

While the filial or second generation of the multigeneration reproduction study and the developmental toxicity study both involve in utero exposure to the embryo and fetus, there are several important differences between the two studies. Among the most important of these differences are the patterns of exposure.

- First, the dose levels in the developmental toxicity study may be much higher than those which are utilized in the multigeneration study.

- Second, as discussed above, the period of treatment in the developmental toxicity study is short, usually encompassing only the period of major organogenesis. Dosing occurs during the critical period of organogenesis (days 6-15 in the rat), and the short duration of exposure may not result in the achievement of a steady state in the fetal/embryonic compartments. If dosing had continued long enough to achieve a steady state, a lower NOEL may have been achieved.

- The third major difference related to exposure is that the route of administration is usually oral (gavage) in the developmental toxicity study versus dietary exposure in the reproduction study. This may result in much higher peak plasma levels of the test material due to more rapid absorption in the study conducted by gavage. The bioavailability of the test material may also be greater via the gavage route of administration than after incorporation in the diet, resulting in a greater area under the plasma curve (AUC) (see Section V.B.6. for further discussion of pharmacokinetics and reproduction indices).

There are other differences between the studies which are not related to exposure. The level of individual fetal examination is generally more extensive in the developmental toxicity study while behavioral effects resulting from in utero exposure may be more amenable to study in the reproduction assay.

Cannibalism of pups with abnormalities, a common maternal rodent response, may reduce the sensitivity of the multigeneration study in the detection of malformations. Because the fetuses are delivered by caesarian section, cannibalism is not a factor in the developmental toxicity study. Given the above, it is not surprising that qualitative and quantitative differences are often observed between multigeneration reproduction studies and developmental toxicity studies.

The dominant lethal test is designed to assess the mutagenic potential of a test substance to the germinal cells of the whole animal and can facilitate the interpretation of two-generation reproduction studies in which treatment-related effects on fertility are observed (see Section II.C.6.). Such information can be useful in determining the site and potential mechanism of action of a test substance.

Information available in humans is potentially the most important ancillary information to be utilized in the assessment of the potential of a chemical to cause reproductive toxicity, since the final objective in the review of toxicity studies is the protection of human health. Clusters or case reports of reproductive toxicity should be considered in light of the toxicity observed in multigeneration studies. Well-conducted epidemiology studies on reproductive toxicants are rarely available for individual pesticides, and even when they are available, exposure is not well-defined. Nevertheless, such studies have provided important confirmatory evidence of male reproductive toxicity in humans for ethylene dibromide and dibromo-chloropropane (Whorton and Milby, 1980). The development of appropriate biomarkers for reproductive toxicity may facilitate future investigations of reproductive toxicity in humans and allow a better correlation of effects observed in animal studies with those observed or predicted in humans (see Table 15A,B, Section V). A recent symposium discussed information for wildlife stuidies which may be useful in indentifying potential hazards for humans (World Wildlife Fund, 1992).

II. DATA ACCEPTABILITY

A. General Requirements

A list of acceptance criteria for §83-4 two-generation reproduction studies is presented in the FIFRA Accelerated Reregistration Phase 3 Technical Guidance (USEPA, 1989b). Intended for use as a checklist for registrants who are submitting study data for reregistration purposes, it can also provide HED reviewers with a concise summary of data which are important to an acceptable

reproduction study. The acceptance criteria are listed below.

An acceptable multigeneration reproduction study should include the following:

1. Technical form of the active ingredient tested.
2. At least 20 males and sufficient females to yield 20 pregnant/dose group.
3. At least 3 dose groups and a control.
4. At the high dose, parental toxicity is observed (or a limit dose is given, 1,000 mg/kg/day).
5.* At the low dose, no reproductive effects are observed.
6.* Analysis for test material stability, homogeneity, and concentration in dosing medium.
7. P animals 8 weeks old at the start of the study.
8. Dosing is continuous starting with the P animals until an individual animal is sacrificed.
9. Mating is 1 male to 1 female.
10. The mating period is not more than 3 weeks.
11. At least two generations are bred.
12. Individual daily observations.
13. Individual body weights.
14. Individual food consumption.
15. Individual litter observations.
16. Individual litter weights (pup weights) at birth and on days 4, 7 (optional), 14, and 21.
17.* Sacrifice schedule, all mating males immediately after last mating, all breeding females immediately after weaning last litter, all animals not used for breeding immediately after weaning.
18.* Necropsy on all animals.
19.* Histopathology of reproductive organs from all animals on the high dose and control P and F1 animals selected for mating. Animals from all other dosing groups if histological effects are observed at the high dose.
20.* Histopathology of all organs with gross lesions.

(Criteria marked with a * are supplemental and may not be required for every study.)

The 1982 USEPA Office of Pesticide Programs, Pesticide Assessment Guidelines "Subdivision F: Hazard Identification: Humans and Domestic Animals", Sections 80-4 and 83-4, define protocol and reporting requirements for

multigeneration reproduction studies. In all cases, scientific judgment must be exercised regarding the deviations of specific study protocols or conduct from published guidelines. The guidelines should not be construed as absolute requirements, and modifications of standard protocols may be appropriate on a case-by-case basis. Significant deviations from the prescribed guidelines require an adequate justification from the testing laboratory. Experts within the Health Effects Division, and elsewhere within the Agency, should be consulted if there is any uncertainty regarding study acceptability.

Some common issues concerning study acceptability are outlined below:

1. Individual animal data must always be provided in submitted study reports. These data should allow for tracing of each parent and its offspring throughout the two generation study. Without such data, it is not possible to determine whether sibling matings have occurred.

2. The highest dose level must induce toxicity (either general or reproductive). This is necesary to maximize the sensitivity of the study. It is preferable that the test chemical not induce mortality in treated animals. The lowest dose level should not induce any treatment-related advers effects in parents or offspring.

3. Although standardization (culling) of litters is suggested in the US EPA Guidelines, it is not a requirement for acceptability. Standardization facilitates study conduct and statistical analysis but may also reduce study sensitivity, resulting in the loss of information (Palmer, 1986). Studies conducted under the recommendations of the 1983 OECD guidelines for reproduction studies will generally not include standardization in the study design, but such a protocol would not jeopardize study adequacy.

4. A one-generation reproduction study is not acceptable to fulfill the data requirement for a multigeneration study. Effects are sometimes observed only in the second generation. Chemically induced <u>in utero</u> effects on development or fertililty are more adequately evaluated in a two-generation study (see Section IV.A). Nevertheless, one-generation studies may provide useful, although supplementary data.

5. Compound administration must be continuous throughout the study. It is preferable that dosing remain constant on a body weight basis throughout the premating, mating, lactation, and weaning periods although in reality, studies are seldom performed in this manner.

6. The study must include an adequate histopathological investigation of the reproductive organs. Histopathological changes are often observed at dose levels lower than those at which fertility and other reproductive indices are affected (see Section IV.A.8.).

7. Studies should comply with Good Laboratory Practice requirements established by either EPA or OECD.

B. Study Design

A standard two-generation reproductive toxicity study, conducted according to the protocol design of the 1982 USEPA Pesticide Assessment Guidelines (§83-4: Reproductive and Fertility Effects), provides information on the reproductive function of the parental animals and the growth and survival of the pups. Study results can provide specific data on germ cells; gametogenesis; libido; fertilization; implantation; gestation; parturition; embryonic, fetal, and neonatal growth and survival; lactation; and postnatal growth and maturation.

1. Standard Protocol

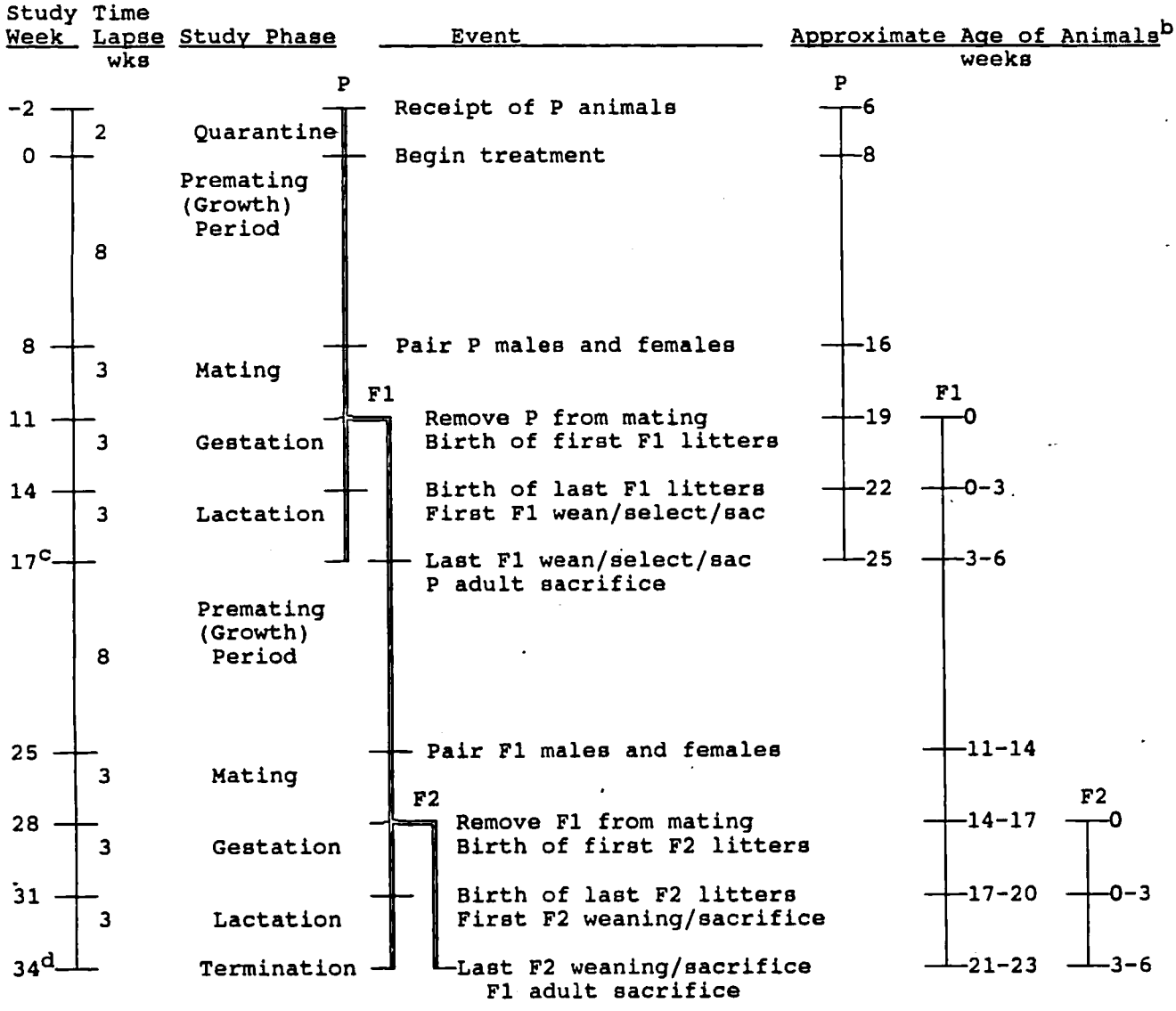
A standard two-generation protocol is diagrammed in Figure 1. In this study, immature rats (approximately 8 weeks of age) are assigned to the first (P) generation and placed in one control and at least three treatment groups. The guidelines specifies that enough animals be assigned to the study to ensure 20 pregnant females per group. The rats are treated daily throughout the period of growth and maturation (approximately 8 weeks) and then mated (1 male : 1 female) for a maximum of 3 weeks. (Treatment continues until sacrifice for all animals.) During mating, the females are checked daily for positive evidence of copulation. Litters are delivered and the F1 pups are raised to weaning (day 21 postpartum) by the dams. At that time, a representative sample of F1 weanlings are selected as parents for the second generation, and the remainder of the study animals are sacrificed. A gross pathological examination is performed on all adult animals, and specified tissues are saved for histopathological evaluation. The selected F1 pups are maintained on treatment until they are approximately 17 weeks old and then paired for mating. The second generation of the study progresses through mating, gestation, lactation, sacrifice, and postmortem procedures in the same manner as the first generation.

Further details on specific aspects of study design and conduct, as well as guidance which can assist the reviewer in determining whether or not a particular study protocol meets guideline requirements, is presented in Section III of this document.

2. Issues Concerning Study Design

a. The One-Generation vs. the Two-Generation Study

Figure 1. Standard Two-Generation Reproductive Toxicity Protocol Design (Rats)a



a Based upon the 1982 FIFRA reproductive toxicity guideline (§83-4). This study design includes only one litter per each of two generations.

b Age ranges for F1 and F2 pups are the result of a potential 3-week variance in conception day.

c Often designated as Week 0 for F1 generation.

d Although the guideline specifies a minimum of 28 weeks for total study duration, this number does not account for variations in the time from pairing to mating among individuals and the resulting overlaps in the period of mating through weaning.

The EPA held a workshop (1987) to discuss whether it was necessary to continue to require a two-generation study in light of a review study by Christian (1986). Table 1 summarizes data presented at the Workshop; it includes three independent comparisons of effects observed in the first and second matings of the first and second generations. The table indicates that toxicity is most often seen at the time of the first mating (first litter) and has been used as justification for modifying the design of the multigeneration study to include only a single mating. It is also of interest that effects are often seen in the first mating of the parental generation although the treatment duration is less than in subsequent generations.

Table 1. Detection of First Effects in Selected Reproduction Studies

Generation	Mating	No. of studies[a]		
		Clegg[b]	HRC[c]	Christian[d]
P	1st	32	23	20
	2nd	5	0	--
F1	1st	5	2	0
	2nd	0	0	--
F2	1st	2	0	0
	2nd	0	0	--
Studies with no effects		27	9	35

a The number of studies in which an effect was first detected in a given generation or mating.
b Author of comparison study (USEPA, 1987)
c HRC = Huntington Research Center.
d Christian (1986)

The Workshop noted that chemicals which bioaccumulate require a second generation, but that adequate pharmacokinetic data which would provide evidence of bioaccumulation are rarely available prior to the initiation of multigeneration reproduction studies. It was generally agreed, however, that it would be beneficial if better information concerning pharmacokinetics were available prior to initiation of reproduction studies as it could, for example, enable one to modify a study design to ensure that plasma levels in animals have achieved a steady state condition prior to mating. Such an approach may be necessary with organochlorine compounds and other compounds with long half-lives for elimination. A workshop sponsored by the USEPA Office of Pesticide Programs (OPP) was held at the National Academy of Sciences (1992) to discuss the use of pharmacokinetic data and risk assessment. A tiered approach to the gathering of pharmacokinetic information was proposed in which the first tier of information

would be developed early in the testing of a compound. The first tier would consist of limited investigation which would yield half-lives for elimination and other basic information. Additional information, such as pharmacokinetic data from pregnant animals, would be generated if triggered by concerns in the area of reproduction or developmental toxicity.

Bioaccumulation is not the only factor which may account for effects seen after the first generation. Effects in parental animals of the second generation may be due to the fact that the F1 animals are exposed in utero, via lactation, and directly from the time of weaning, whereas exposure of the P generation is not begun until the animals are 6 to 8 weeks of age. The ability of perinatal rodents to metabolize agents is limited due to a lower level of mixed function oxidase and conjugation capacity compared to adult animals. It is not until the animals are 30 to 60 days of age that metabolism and excretion capabilities approach that of an adult (USEPA, 1987).

In summary, although most studies of reproductive toxicants will demonstrate reproductive effects after the first mating, some chemicals will show effects at lower levels, or qualitatively different effects, after the second mating due to exposure during a longer portion of the lifespan, and, for chemicals with long half lives for elimination, more time to achieve a steady state level. For these reasons, a second generation is required in the standard protocol for a reproduction study.

b. The Need for a Second Litter in Each Generation

Bioaccumulation is also relevant to the need for a second litter in each generation. Chemicals with very long half lives for elimination (such as DDT) may not reach a steady state concentration at the time of mating for the first litter and the second litter may be more sensitive for the detection of reproductive effects for these chemicals. The duration of treatment of the parental animals is also longer at the time of the second mating, and the animals are older, larger, and sexually experienced. All of these factors may lead to the second litter being more sensitive than the first litter. However, the need for a second litter in multigeneration studies should be considered on a case-by-case basis and should not be required without justification. A second litter was routinely included in studies conducted prior to the mid-1980s.

c. Length of the Premating Period

The length of the premating treatment period has ranged in various

protocols from 8 to 14 weeks for rats. The 1982 FIFRA test guidelines suggest, but do not require, a premating period of 14 weeks. The OECD and FDA test guidelines for a reproduction study recommend a mininum of 10 weeks. The premating period must be at least long enough to cover the duration of spermatogenesis plus epididymal transit time in males and a period of 5 estrous cycles for females. Because the female rat cycle is only 4 or 5 days, the limiting factor in the necessary dosing period prior to mating is ensuring test compound exposure throughout all periods of spermatogenesis. In some cases, pregnancy rate may may be related to the length of the premating period, with longer premating periods resulting in a lower and more highly variable fertility (Palmer, 1981). A short premating period may result in younger animals of low body weight which have low birth weight pups. Although the optimum period will vary slightly among strains of rat, the 10 week minimum period recommended in the OECD test guidelines should be considered as sufficient for commonly used strains of rats.

d. Standardization of Litters

The 1982 FIFRA test guidelines recommend standardization (culling) of litters to 4 males and 4 females on day 4 postpartum. As noted in Table 2 (Section II.B.), OECD test guidelines make the culling of litters optional. The FIFRA Accelerated Reregistration Phase 3 Technical Guidance (1989) indicates that standardization of litters, as described in the FIFRA guidelines (83-4) should not be considered as mandatory.

It should be noted that many reproductive toxicologists prefer standardization of litters. Variability is decreased and consistency in litter size facilitates statistical analysis. Unusually large (or small) litters may affect mean pup weight and survival and may complicate the identification of compound-related toxicity on these parameters. In addition, it has been stated that study sensitivity actually increases with standardization. This latter assertion has not been tested experimentally or statistically.

Some arguments for not standardizing litters have been advanced by Palmer (1986). Among the list of issues raised are the following:

1. Standardization disrupts the normal distribution of litter sizes.
2. The standardized litter size is below the natural mean, median, and modal values which are normally observed.
3. Standardization results in the alteration of, on average, 77 percent of normal litters in young rats, and 91 percent of litters in older

rats.

4. Standardization results in the elimination of 25 to 40 percent of normal offspring, most of which would survive.
5. Human bias or error is introduced in the selection of retained animals.
6. The introduction of human bias preferentially acts against controls where standardization is less likely to alter the mean, median and modal values.
7. Standardization raises mean body weight at weaning and thus reduces the likelihood of discovering a lactational effect on body weight.

Mean pup weight and mean litter size are the two factors which determine the mean litter weight. The mean litter weight is most analogous to neonatal weight in monotocous species such as the human. Effects on mean pup weight and mean litter weight may not be as readily detected after standardization due to the loss of information which would have been obtained from the culled pups.

C. Acceptability of Other Protocols

Some basic comparisons between the EPA study design and other commonly followed protocols are shown in Table 2.

The differences between these protocols are relatively minor and should not present a barrier to study acceptance. Differences between OECD and Agency guidelines are not major and should not affect study acceptability. In addition to the protocols shown in the following table, the OECD SIDS protocol, the Chernoff-Kavlock screen, and a one-generation reproduction study may provide useful supplementary information (see Section II.C.5.). The latter is particularly helpful in assessing dose selection in the multigeneration study.

The FDA multigeneration study protocol referenced in Table 2 is described in more detail below. In addition, FDA Segments I, II, and III protocols, Reproductive Assessment by Continuous Breeding (RACB), and the dominant lethal test are discussed. While of potentially great value in identification of a reproductive toxicant, the more limited scope of these various study protocol generally preclude their use alone for risk assessment purposes (see discussion of use of ancillary reproduction studies in risk assessment, See Section V. B. 2.).

1. FDA Multigeneration Study

The initial three generation reproduction study was developed by the US Food and Drug Administration for the testing of intentional food additives, i.e., food coloring agents, sweeteners, and unintentional food additives including pesticides (USFDA, 1959). The protocol was modified in the early 1960s and was recommended by the President's Science Advisory Committee (1963) for the testing of pesticides. It has evolved into the multigeneration reproduction study which is required by the Office of Pesticide Programs. The original protocol called for treatment to be initiated in mature male rats 60 to 80 days prior to

Table 2. Comparison Between Multigeneration Study Designs Recommended by US and Foreign Regulatory Agencies

Protocol by Agency[a]	EPA[b]	FDA	OECD
No./sex/group	20	30 for P 25 for F1	20
Age at start of dosing (P)	8 weeks	After weaning	5 to 9 weeks
Premating period (P)	8 weeks	10 weeks	10 weeks
Mating ratio (M:F)	1:1	1:1	1:1 or 1:2
Standardization (culling)	Yes[c] (4 males, 4 females)	Yes (5 males, 5 females)	Optional (4 males, 4 females)
Disposition of P animals	P males after mating, females after weaning	P males after mating, females after weaning	P males after mating, females after weaning
Parental selection (F1)	1 male and 1 female/litter; no sibling matings	2 males & 2 females/litter	1 male & 1 female/litter; no sibling matings
F1 premating period	14 weeks postweaning	13 weeks postweaning	10 weeks postweaning

a Criteria listed are for rats.
b EPA FIFRA guidelines only; TSCA protocol (USEPA, 1985a) differs as follows: 1) 5-8 weeks of age for dosing (P), 2) premating period (P) is 10 weeks, 3) F1 parental selection, change mate after 1 week, 4) F1 premating period is 11 weeks postweaning
c Interpreted as optional.

mating (the time required for spermatogenesis and epididymal transport of mature sperm) while females were to be treated for 14 days prior to mating (the length of three standard estrous cycles). Daily vaginal inspections were made with the finding of sperm or a copulation plug considered as day 0 of pregnancy. Each generation was required to produce two litters with the first litter discarded at weaning. Two males were originally paired with each female (this was later revised to a 1:1 mating ratio), and extra young and adults were sacrificed.

After the young from the second mating of the final generation are weaned, histopathological examinations and organ weight analyses of 10 male and 10 female offspring per group were performed.

2. The FDA Segment I, II, and III Studies

The FDA Segment I, II and III screen (Goldenthal, 1966) was originally proposed to replace the two-litter rat reproduction study for the appraisal of new drugs for use during pregnancy and in women of childbearing potential. These segments are: I. Study of Fertility and General Reproductive Performance; II. Teratological Study; and III. Perinatal and Postnatal Study. The Segment I, II, and III studies, either alone or in combination, cannot be used to satisfy EPA regulatory requirements for a multigeneration study although they may be used in the process of risk assessment (See Section V. B. 2. on the limitations of non-multigeneration reproductive toxicity protocols).

a. Segment I

In Segment I, mature male rats are treated with the test compound for 60-80 days prior to mating. Female rats are treated for 14 days, then mated to the treated males. Daily vaginal inspections are made, and the finding of sperm or a copulatory plug is considered as day 0 of pregnancy. Dosing is continued throughout pregnancy. One-half of the females are sacrificed on day 13 and examined for number and distribution of live and dead embryos. The remaining dams are allowed to litter normally, and the newborn pups are counted, examined, sexed and weighed. The pups are weighed and counted again on days 4 and 21 (Collins, 1978b).

b. Segment II

Segment II studies investigate the effects of a chemical on <u>in utero</u> development. They are equivalent to the protocols for the rodent developmental toxicity study required under FIFRA.

c. Segment III

The Segment III perinatal and postnatal study is used to evaluate the effects of drugs on late fetal development, labor and delivery, lactation, and newborn viability and growth (Collins, 1978a). It begins with drug administration during the final one-third of gestation and continues through lactation to weaning. Observations are made for effects on parameters such as

labor and delivery, the duration of gestation, litter size, and pup weight. Continuous compound administration through lactation allows for detection of adverse effects on lactation ability as well as any toxic reactions due to the drug or metabolites on the newborn as a result of excretion in the milk. Some of the offspring may continue to be observed to adulthood.

3. Continuous Breeding Protocol

The continuous breeding protocol was developed by the National Toxicology Program as a possible alternative to the multigeneration study and has been evaluated using both mouse and rat models (Lamb and Chapin, 1985; Morrissey et al., 1989, Gulati, 1991). The study allows for continuous breeding of treated males and females for 14 weeks (after pretreating for 7 days) with immediate removal of each litter after birth. The exception to this is that the final litter may remain with the dam until weaning or may even remain on study to maturity with subsequent mating. Mating of treated males with untreated females, or treated females with untreated males, can sometimes allow identification of gender-specific effects.

This protocol allows for up to five litters per pair during the 14 week study. It also may allow for determination of the affected sex by mating treated animals with untreated controls (crossovers) for 7 days (with treatment discontinued for this period). This protocol offers considerable advantage. It may be more sensitive due to the fact that the study design results in a larger number of matings and subsequent pregnancies which allows a better assessment of the onset of effects on fertility. The continuous mating could be incorporated into the multigeneration study. The second generation (F2) may be taken from the first mating, rather than the last, further shortening the length of the study. Alternatively, the F1 offspring could be used for the continuous breeding segment. Modifications of the continuous breeding protocol may be acceptable to the Agency on a case-by-case basis. Registrants should be encouraged to discuss individual study designs which are based on the continuous breeding protocol with HED scientists prior to study initiation.

4. Combined Reproduction and Developmental Toxicity Study

A modified multigeneration study which includes a developmental toxicity phase has occasionally been conducted for pesticides, usually with fetal visceral and skeletal examinations performed on the second litter from the second (or third) generation. Although this study design may be acceptable for the purpose of assessing reproductive toxicity, it is rarely an adequate assessment of

developmental toxicity potential. The dietary exposure levels of the reproduction study are generally inadequate to achieve the degree of maternal toxicity required in a developmental toxicity study. The combined reproduction/ developmental toxicity study will not be discussed further in this SEP.

5. Screening Protocols

In the attempt to reduce the significant time and cost burdens incurred in conducting a typical two-generation protocol, screening protocols have been established to detect developmental (Chernoff and Kavlock, 1982) and/or reproductive effects (Gray et al., 1988; Harris et al., 1992) in rodents.

Through the OECD, the 1992 Screening Information Data Sets (SIDS) has been developed with the intent of providing initial information on possible reproduction and/or developmental effects at an early stage of assessing the toxicological properties of a substance, on chemicals of high concern, or on existing chemicals for which there is little or no toxicological information available. Although negative data do not provide evidence for definitive claims of no effect and do not indicate absolute safety with respect to reproduction and development, positive findings are useful for initial hazard assessment and contribute information relative to requirements for additional testing. The protocols, although not yet finalized, are being used both in Europe and in the U.S. (under the Toxic Substances Control Act). In the study design for the Guideline 421 Reproduction/Developmental Toxicity Screening Test, 10 rats/sex are assigned to 3 treated and 1 control group. They are dosed daily with the test substance by gavage for 2 weeks prior to mating and until sacrifice. The animals are mated (1:1 for a maximum duration of two weeks) and the females are allowed to deliver. Litters are counted, sexed, weighed on days 1 and 4 postpartum, and sacrificed (along with the dam). Males are killed at least 14 days, and unmated females at 24-26 days, after the last day of the mating period. Minimum postmortem procedures include external examination of the pups; macroscopic gross pathology of the adults; implantation site counts; testes and epididymides weights; and microscopic evaluation of ovaries, testes, accessory sex organs, and all organs with gross lesions for control and high dose animals. Reviewers should be aware that screening protocols, including the SIDS protocol, do not meet the requirements for a §83-4 two-generation study in rodents under FIFRA, and would be considered CORE-Supplementary studies.

6. Dominant Lethal

Dominant lethal studies are conducted in either rats or mice (Green et al.,

1985, 1987). Only one sex is treated, generally the males. Following acute or subacute treatment of the males with the test substance in a standard dominant lethal protocol, they are mated (1:1 for a maximum of 5 days) each week with untreated virgin females for a total of 10 weeks in rats (8 in mice). The weekly mating schedule allows evaluation of the effect of treatment on various stages of spermatogenesis. The females are sacrificed at midgestation (approximately days 11-17 postmating), and their uteri are examined for live and total implantations and early and late deaths. An increase in embryo deaths is indicative of a treatment-related effect on spermatogenesis, and adverse effects on specific spermatogenic cell type(s) can be identified.

III. EVALUATION OF STUDY CONDUCT

A. Choice of Test Compound

Section 83-4 of the 1982 FIFRA Guidelines states that a multigeneration test shall be performed with the technical grade of each active ingredient in the product. The test compound selected should be the technical product intended for commercial use. In some cases testing is conducted using a material which is produced before the commercial manufacture of the technical material, i.e., before manufacturing processes are "on-line". The specifications of the test material should be available, and the concentration of the active ingredient(s) and of the impurities should be clearly indicated in the study report. This information should also be included in the study evaluation for comparison with material utilized in future studies. If a vehicle is used, it should not produce any systemic or reproductive toxicity. If there is a question as to the toxicity of the vehicle, the registrant should be required to justify the choice of the vehicle.

B. Animal Selection

1. Species Selection

Criteria in the selection of the test species in a multigeneration reproduction study include animal size, length of gestation, litter size, fertility rate, ovulation, ease of maintenance, and comparability of metabolism of the chemical with that in human (Collins, 1978b). Of the species available, the rat and mouse are preferred due to their small size, short gestation time, high fertility rate, spontaneous ovulation, short estrous cycle, and ease of maintenance (see Table 3). Golden Syrian hamsters, although having a short gestation time (16 days) and a large litter size (6-10 pups), are not often used

Table 3. Some Parameters of Reproduction[a]

Event	Rat	Mouse	Rabbit	Guinea Pig	Hamster	Gerbil	Ferret[b]
Male breeding age	100 days	50 days	6-7 mo	3-5 mo	2 mo	10-12 wk	8-12 mo
Female breeding age	100 days	35-60 days	5-6 mo	1-5 mo	2 mo	10-12 wk	8-12 mo
Type of estrous cycle	Poly.	Poly.	Poly.	Poly.	Poly.	Poly.	Mono. (Mar-Aug)
Length of cycle (days)	5	4-5	Cont.	16.5	4-5	4-6	Cont.
Duration of estrus	10-20 hr	10 hr	Cont.	6-11 days	20 hr	1.4	Cont.
Time of ovulation (hours after onset of estrus)	10	2-3	10-11 p.c.	10	8-12	6-10	30-40 p.c.
Gestation length (days[mean])	20-22[21]	17-21[19]	30-32[31]	65-71[68]	15-17[16]	24-26	41-43
Litter size (range[mean#])	8-12	[11]	1-18[8]	1-6[4]	6-10	1-12(4.5)	1-12(8)
Birth weight (g)	5-6	1.5	60	90	2	2.5-3.5	6-12
Weaning age (wks)	3	3	8	3.5	3	3	6
Weaning weight (g)	40-50	7-15	1500	250	35	- -	- -
Duration of ability to reproduce (years)	1	6-10 litters (♀) 1.5 (♂)	1-3	4-5 (♀) >5 (♂)	1 (♀) 2 (♂)	2	2-5

a Adapted from Hafez (1970), except as noted.
b From Fox, 1988.
Poly. = Polyestrous
Mono. = Monoestrous
Cont. = Continuous estrous cycle
p.c. = Postcoital

in a reproduction study due to their variable or unpredictable fertility rate. Mongolian Gerbils have proven to be useful in a reproduction study (Robinson, 1979) due to their size, short gestation period, ease of maintenance, spontaneous ovulation, and fertility rate. However, gerbils are less prolific than rats or mice, since their litter size averages 4-5 pups. In addition, they tend to bemonogamous in captivity (Holmes, 1985), although this may be an artifact of housing techniques (Thiessen and Yahr, 1977). Rabbits are rarely used in a reproduction study since they do not ovulate spontaneously and have a relatively long gestation period (31 days). Further, rabbit maintenance is expensive. A long gestation period (65-71 days), coupled with a small litter size (1-6 pups), makes guinea pigs unsuitable for reproductive studies. Primates, dogs, and pigs generally do not provide any major advantage over rats or mice and are not recommended for routine testing. It has often been stated that the choice of the test animals should be based on similarities to humans with respect to such characteristics as plasma concentration, placental transfer, metabolic patterns, and embryonic developmental schedules. However, there is no species which closely resembles humans for all of these parameters that can be used for routine testing. Therefore, for reasons of convenience and the availability of an extensive historical data base, the rat is the preferred species for a reproduction study. An exception is the Fischer 344 strain of rat, which has proven to be an unreliable model for reproduction studies. Further discussions are limited to the rat.

2. Health Status

Virgin animals should be used. Upon arrival at the testing facility, all animals must be quarantined for at least 14 days, during which time their health status must be checked by a veterinarian. Quarantine of newly arrived rodents reduces the possibility of transmitting active infections and allows animals to adjust to their new surroundings.

At study initiation, all animals must be sexually mature (approximately 8-10 weeks old) and disease-free, since treatment or infections during the course of the investigation may lead to unpredictable pharmacological or toxicological outcomes.

3. Number of Animals

Reproduction studies customarily require 20 pregnant female rats (or mice) per dosage level; therefore, a study starting with exactly 20 females per group risks rejection by the Agency reviewer if an inadequate number of litters is

obtained in any group. At least 20 pregnant females at or near term are needed as parental animals for each dose group; numbers fewer than this may reduce study sensitivity. The investigators must ensure that an adequate number of litters are obtained for each control and treatment group and this may require mating 30 or more females to achieve a sufficient number of litters at term. Studies with fewer litters than 20 per dose level may be considered acceptable (CORE minimum; see Section IV.D.) on a case-by-case basis. Studies with significantly less than 20 litters at dose levels other than the high dose are generally considered to be inadequate for the purpose of meeting regulatory requirements. The reviewer can use the criteria listed below in determining the acceptability of studies that do not meet the numerical litter requirements.

The following criteria should be met in order to ensure that the data was statistically analyzed with a high degree of confidence and that the study can contribute meaningful information for risk assessment analyses:

a. The number of control litters should be $\geq$20. If the control groups in both generations present evidence of a fertility problem, the genetic vigor of the rats procured for the study should be questioned by the reviewer. The fertility data from such a study (at all dose levels) would be of limited reliability.

b. The study must provide enough information to determine the NOEL.

c. The number of litters should be $\geq$20 at the dose level which is determined to be the NOEL.

Since it is not possible to describe in detail every possible combination of factors that the reviewer may encounter when confronted with a study having one or more groups with less than the recommended 20 litters, the following philosophies may also aid the reviewer in judging the acceptability of such a study.

a. The most important criterion, and the one upon which all decisions of study acceptability must be made is this: that there be enough information provided in the study data to adequately assess the effects of the test substance on the reproductive performance of the adult animals and on the development, birth, growth, and survival of the offspring through lactation. For example, if a reduced number of litters is observed sporadically, i.e., in one or two dose groups (not dose-related) or in one generation only, a judgement must be

made as to how much information has been lost and whether or not this information gap might compromise the interpretation of study results.

b. Enough litters should be produced in the first generation to provide an adequate number of weanlings for selection as second generation parental animals, allowing for sufficient genetic variability and avoidance of sibling matings. Achieving the recommended 20 litters in the second generation is not as critical.

c. The doses on a toxicology study are generally selected to produce toxic effects in the highest dose group; it is not unusual for such toxicity to interfere with the reproductive process. Excessive treatment-related toxicity, resulting in a reduced number of litters in the high-dose, and to some extent in the mid-dose group(s), may not compromise the acceptability of the study if the study NOEL can still be determined, even if the high dose group is not able to continue into the second generation.

d. Additional information provided by the registrant may be used to increase the sample size (number of litters) in deficient groups. The reviewer may find that the study includes data from an additional mating and litters, e.g., F1b and F2b, in each generation. These additional matings, although not required, are often conducted if less than 20 litters were produced in one or more groups from the first mating. Alternatively, the registrant may submit other supplementary reproduction study(ies) for review and consideration. Because the animal populations and methods may not be exactly equivalent between the studies or study segments, the reviewer must scrutinize the data thoroughly to identify any differences and determine whether those differences can interfere with the interpretation of the study results.

C. Environmental Factors

Environmental factors can have an important influence on the toxicity of a test compound and should be described in each study report. Environmental factors which may influence the reproductive performance of the test animals or exert unnecessary maternal stress include the following:

1. Temperature and Humidity

Adaptations to changes in ambient temperature in rodents consist of peripheral vascular constriction, piloerection, increased metabolic activity which may result in increased food consumption, and variations in biologic and physiologic effects in response to fluctuations in the ambient temperature. Daily fluctuations in room temperature and humidity may act as significant stressors (Kohn and Barthold, 1984). Yagil et al. (1976) reported that the production of milk was impaired in rats exposed to 35°C for 8 hours/daily. Temperatures of 32°C or greater have also been demonstrated to impair the reproductive capacity of rats (Yamauchi et al., 1981). Room temperatures between 72° and 76°F (22-24°C) are desirable, and the humidity should range between 40 and 60%. It should be noted that high ambient temperature and humidity may increase the susceptibility of the test animals to infectious agents (Baetjer, 1968) and may cause male infertility (Baker, 1979).

2. Light Intensity and Timing

Seventy five to 125 fc (foot-candle) are suggested as an optimal range for light intensity which should be evenly distributed to all animals in the room. However, there is evidence of retinal degeneration in laboratory animals at that recommended light intensity range (Anver and Cohen, 1979; Bellhorn, 1980). Light is a stimulant and synchronizer of the reproductive system (Pakes et al., 1984) and is thought to be mediated through the hypothalamus (Wiehe, 1976). Consequently, photoperiod (light-dark cycle) not only can modify the biologic response but also has a profound effect on the circadian rhythms of rodents (Hastings and Menaker, 1976). For a reproduction study, a daily light period of 12-14 hours appears to be optimum (Mulder, 1971). Standardization of light intensity and duration, accomplished by the use of cage rotation techniques and automated light timing equipment in the animal room, is necessary in a reproduction study since the cyclicity of estrus and ovulation in rodents is controlled by the diurnal rhythm of the photoperiod. Constant light for as few as 3 days may induce persistent estrus and polycystic ovaries (Baker et al., 1979).

3. Nutrient Requirements

The final report for any reproduction study must contain information concerning the type and frequency of diet analyses as well as identification of any contaminants that were found. Nutritionally adequate rodent diets are readily available commercially, and the manufacturer generally provides the

purchasing laboratory with, at the very least, nutritional content analysis data. Generally the study report will not contain these data unless the report was conducted under Japanese guidelines or a problem with nutritional content of the diet was identified.

Nutritional deficiencies are generally not a problem with contemporary studies conducted in major testing laboratories; however, the reviewer should be aware of possible effects of deficiencies that are identified in study reports. The presence of a nutritional deficiency may result in adverse reproductive outcomes such as (1) irregular cyclicity, (2) delayed puberty, (3) longer time intervals necessary to become pregnant, and (4) loss of total reproductive lifespan (Ball et al., 1947; Berg, 1965; Frisch, 1978; Merry and Holehan, 1979; Nelson and Felicio, 1984). Nutritional requirements vary with the physiologic condition of the animals. Test animals usually require a higher intake of proteins, minerals, and vitamins during gestation and lactation. For example, the minimum concentration of protein needed for maintenance of adult rats is 4.5% of the diet but increases to 12.0% during gestation and lactation (NRC, 1978). Vitamin deficiencies (E, riboflavin, thiamin) may be associated with infertility. Nutrition-related alterations in reproductive performance may be reversible, however, once the causative factor has been corrected.

Commercially available diets are formulated from natural products and thus are subjected to changes in nutrient composition. Batch-to-batch differences exist in commercial food. For that reason, it is preferable that the lot or batch of commercial chow used in the preparation of the diet throughout the entire investigation should remain the same. It should be possible to verify this information from the study report.

The reviewer must also be aware that contaminants may be identified in the diet analyses, and that such contaminents may have pronounced effects during gestation and lactation of rodents. Some examples include low levels of heavy metals, insecticides, mycotoxins, or synthetic estrogens. These unintentional contaminants may occur naturally in plant materials or remain as residues from agricultural pesticide uses. Nitrosamine may be found in diets using fish meal as a major source of proteins, and aflatoxins may be detected in corn, wheat, and other cereals during storage. Many of these contaminants have resulted in serious effects on reproduction. They may act as developmental toxicants (lead, mercury, cadmium, aflatoxins) or may prevent the implantation process as well as delay fetal growth (lead, cadmium; Degraeve, 1981).

4. Drinking Water

Drinking water analyses are conducted for/by most major testing laboratories and may include information regarding contamination at the source of the water (i.e., the municipal water treatment plant), at a distribution point, or from the specific animal room. The water may be routinely analyzed for such specifics as ionic content, heavy metals, pesticide residues, or bacterial contamination. Contaminants may be classified as suspended solids, organic solutes, or inorganic solutes (Shapiro, 1980). The suspended solids are mostly harmless; however, the organic solutes such as cyclic aromatics and halogenated hydrocarbons may exert an effect on the physiological response of the test animals. Organic contaminants which are frequently found in drinking water include halomethanes. These compounds derive from the interaction of a halogen (chlorine, bromide) and methane (from organic materials). The toxicological effects of many of these compounds, such as chloroform, have been well documented (IARC, 1990). Others have not been well investigated. Nitrate is another frequent contaminant of drinking water. Excessive levels of nitrates in water have been associated with methemoglobinemia.

The specific results of drinking water analyses are seldom provided in study reports conducted under the 1982 FIFRA guidelines. However, any unusual findings will generally be discussed in the report; possible effects of contaminants on the study results should be considered by the reviewer.

5. Animal Housing

In a reproduction study, females are usually housed individually in solid-bottom cages except during mating. The type of bedding material should be reported since it may influence the biologic response of the animals. Wood shavings or chips are commonly used; however, hardwoods are preferred to softwoods since aromatic softwoods are well known hepatic microsomal enzyme inducers (Baker et al., 1979). Nesting material (cotton, shredded paper) is usually not necessary for rats.

6. Other Factors

The use of pesticides around the testing area is not recommended. Although sanitation of the testing area is crucial, pesticidal use may confound interpretation of the study results. Pesticides, air-deodorizing agents, and solvents may stimulate or inhibit the microsomal enzymes depending on the chemical used. Room deodorizers which consist of volatile hydrocarbons and

essential oils may stimulate or inhibit the enzymes. Cleaning agents, solvents, and surfactants may have a similar effect. One disinfectant which is commonly used in the laboratory is ammonia. The reviewer should be aware that ammonia is an inhibitor of hepatic microsomal enzymes (Vessel et al., 1976). It is therefore suggested that non-chemical means of sanitation should be used. However, if the use of a solvent or insecticide is unavoidable, it should be clearly stated in the final report.

Antibiotics are sometimes needed to control infectious diseases during an on-going investigation. Antimicrobial agents may also have an impact on the physiologic response of animals. The use of any chemical in a reproduction study must be documented by the investigators, and its potential impact on the study results should be carefully considered by the reviewer.

D. Route of Administration

In the testing of pesticides, the route of administration is generally through incorporation into the diet. Under special circumstances, dependent primarily upon the stability of the test material, its physical and chemical properties, and its most likely route of exposure to humans, it may be given by gavage, inhalation, dermally, or in drinking water.

Dietary exposure is the easiest route of administration since the animals do not have to be handled daily, and the amount of food consumed is directly proportional to the size and metabolism of individual animals. Dietary administration, however, is not appropriate for compounds which cannot be homogeneously mixed into the diet or which degrade rapidly at room temperature. Therefore, analytical determination of the test substance in the diet must be performed periodically and appended to the final report. The reviewer should consider data which describe homogeneity of the diet mixes, stability of the test material in the diet under storage and/or animal room conditions, and the frequency of diet preparation.

Gavage may be the recommended route of exposure if the oral route is indicated but administration through feed or water is neither practical nor appropriate for a particular substance. The nature of the vehicle and additives (suspension and wetting agents) must be indicated. The vehicle and additives used should not interfere with absorption of the test substance or produce toxic effects (USEPA, 1982). Stability data of the test compound in the vehicle should be known. For insoluble test substances that are suspended in a vehicle, the reviewer should ensure that efforts were made to maintain homogeneity of the

suspension during dosing.

Gavage dosing is commonly accomplished at a constant volume of 5 ml/kg of body weight, although larger volumes may be required for compounds with low solubility (Kimmel and Price, 1990). The vehicle control group should receive the same treatment as the groups administered test compound. The dosing volume for each animal must be adjusted to individual body weight. In rodents, weekly adjustments are sufficient. The time of dosing is also of importance in a gavage study. Rodents are nocturnal animals and as such have their peak food consumption prior to day light. Administration of a test compound in the morning hours may result in a decrease in test material absorption due to the presence of food in the stomach. It is suggested that dosing by gavage is preferably performed after mid-morning (Stevens and Gallo, 1989). Handling of animals during the dosing procedure constitutes an additional stress factor to pregnant animals and may lead to resorptions and/or abortions.

E. Dose Levels and Dose Selection

Current test guidelines recommend that at least three dose levels and a concurrent control be used. Subdivision F of the FIFRA Guidelines (USEPA, 1982) clearly indicates that:

1) The lowest dose level should not produce any evidence of toxicity.
2) The intermediate dose level should produce minimal observable toxic effects.
3) The highest dose level should produce some indication of maternal or adult toxicity.

While not currently required, it is preferable that information on metabolism, pharmacokinetics, bioavailability, and/or bioaccumulation of the test substance should be available to demonstrate adequacy of the dosing regimen. It is further recommended that steady state be reached prior to initiation of the mating period.

Among the objectives of a multigeneration reproduction study are the demonstration of a No-Observed-Effect Level (NOEL) and a Lowest-Observed-Effect Level (LOEL) for the parameters which are investigated. The induction of minimal toxicity by the intermediate dose level is desirable but should not be viewed as essential. A LOEL is necessary to indicate that a sufficiently high dose level has been used to elucidate the potential of a chemical to cause reproductive effects. However, in the case that toxicity is not demonstrated with the dose

levels selected, the study may still be considered as acceptable if the highest dose level is at the limit dose (1000 mg/kg/day). Preferably, all dose levels used in a reproduction study should be reported on a mg/kg/day basis. In those cases where dose levels are provided only on a ppm basis, the reviewer should convert the dietary concentrations to mg/kg of body weight (see Section IV.A.4.) using actual food consumption measurements or, alternatively, using the nomimal conversion factors reported in the Lehman tables (Table 4).

Table 4. Approximate Relation of Ppm in Diet to Mg/Kg/Day[c]

Animal	Weight (kg)	Food [a] Consumed per Day (g) (Liquids Omitted)	1 ppm in Food Equals, in Mg/Kg/Day	1 Mg/Kg/Day Equals, in ppm of Diet
Mouse	0.02	3	0.150	7
Rat (young)	0.10	10	0.100	10
Rat (older)	0.40	20	0.050	20
Guinea pig	0.75	30	0.040	25
Rabbit	2.0	60	0.030	33
Dog	10.0	250	0.025	40
		750 [b]	0.075	13
Monkey	5.0	250 [b]	0.050	20
Man	60.0	1500 [b]	0.025	40

a Dry laboratory chow diets, unless otherwise noted.
b Moist semi-soft diets.
c Adapted from USFDA (1959).

F. Exposure Period

The test compound is ideally given to the test animals on a daily, seven-days per week. It is suggested that:

1) Parental (P) generation males and females should be exposed to the compound at a minimum age of 6 weeks, and exposure should continue for at least 8 weeks prior to mating.
2) Female parental animals should be exposed to the test compound during gestation, lactation, the time interval during one or two successive matings (Fa and Fb), and until final sacrifice.
3) Offspring should be exposed without interruption from <u>in utero</u>, through lactation, weaning, and the growth period, and until sacrifice at weaning. Those selected to be parental animals of the next generation should be exposed through mating, the reproduction period, and until final sacrifice. The exposure period from weaning to mating should be at least 8 weeks.

Parental animals and their offspring should be continuously exposed to the test compound. Parental animals, especially males, in some studies which include

two litters per generation, may be placed on control diet during the resting period (the interval period between Fa and Fb); however, this should be considered a study deficiency since all phases of spermatogenesis will not have been exposed to the test compound in the second litter.

G. Mating Procedure

Paired mating (one male to one female) is strongly preferred to colony mating in reproduction studies. The 1983 OECD guidelines recommend mating one male to two females; although this procedure is considered to be acceptable, the reviewer should be aware that statistical analysis of the data must be adjusted, and the HED statistician consulted with any questions. Sibling matings must be avoided. The age of the animals at mating should be checked by the reviewer to ensure that the animals have reached sexual maturity.

Mating is usually confirmed by the presence of a copulatory plug retained vaginally or located on the cage tray and/or the presence of spermatozoa in the vaginal smear. In the rat, a vaginal smear examination for the detection of sperm is more reliable than the presence of a copulatory plug. The 1982 FIFRA guidelines indicate that unmated pairs may be remated with other proven sires or dams of the same group (See Section IV.C.4.b. for further discussion).

The individual animal data should allow the identification of the sire assigned to each dam during the first (or second) mating trials. The day of confirmed mating and delivery for each dam should be reported.

Both the OPP (1982) and OECD (1983) guidelines indicate that males and females should be cohabited until pregnancy occurs or until 3 weeks have elapsed. However, successful mating occurs within 4 days in approximately 90% of all pairings. If pregnancy does not occur in the alloted time, possible causes of infertility in the pair should be considered. Information contributing to this analysis might include results of additional matings, female cyclicity data, sperm evaluations, or histopathological examinations of reproductive organs.

H. Statistical Analysis

The statistical methods used must be described, referenced, and identified, since interpretation of reproductive results as well as of any toxicology data should rest on a sound statistical basis (see Table 5). The analysis of data arising in the reproduction study is complicated by the interdependence of

Table 5. Data Characteristics and Commonly Used Statistical Methods for Hypothesis Testing in Reproductive Toxicity Studies[a]

Type of Data	Frequency Distribution	End Points	Commonly Used Statistical Tests
Continuous	Normal	Body weight Body weight change Food consumption Organ weight[b] Survival rate[c] Pup/litter weight Crown-rump length	Bartlett's test for homogeneity of variance[f] F test[g] Analysis of variance (ANOVA)[h] t test[h]
Discontinuous (nonparametric)	Not normal	Behavioral signs[d] Corpora lutea Implantation sites Live/dead pups	Wilcoxon-Mann-Whitney U test[i] Kruskall-Wallis test[i]
Categorical	Not normal	Behavioral signs[d] Clinical signs Pup alterations[e] Dose/mortality data	Chi square test[j] Fisher's Exact test

a Adapted from Gad and Weil, 1986.
b Absolute and relative weights.
c End points are expressed as proportions or percentages of each litter with pre- or post-implantation losses, early or late resorptions, and live or dead pups.
d Behavioral signs may be discontinuous or categorical.
e Group incidences (number of litters with affected pups/number examined in each group) are compared.
f When Bartlett's test indicates that variances are not homogeneous, data are sometimes transformed by using the square root, the arcsin of the square root or other transformations of each value to stabilize variances before further analyses are attempted. Because some of the statistical tests have become standard practice, Bartlett's test may not be used.
g The F test is the same type of test as Bartlett's except that it tests for homogeneity of variances for just two groups.
h ANOVA is used for comparison of three or more groups of data with homogeneous variances and a normal frequency distribution. The t test compares two groups of continuous normally distributed data.
i If $n \geq 40$, a 2 by 2 chi square test with continuity correction can be done.
j For $n \leq 9$ nonparametric methods should be used because normality and homogeneity of variance are decreased. The Wilcoxon-Mann-Whitney U and Kruskal-Wallis Tests are nonparametric tests that make no mathematical assumptions about the distribution of the data. If $n \geq 10$, a t test is just as useful.

various reproductive parameters and what has sometimes been referred to as "the litter effect". The latter is the result of the lack of independence of various observations observed in the same litter. The similarity of findings which is often observed among litter mates may be due to a variety of causes including genetic similarities, a common maternal environment, and differences in handling between dams. Techniques which have been developed to minimize "the litter effect" are discussed in Khera et al. (1989). Unless effects are clearly related to the male, the female should be used as unit for statistical purposes.

Statistical analyses performed in the Food and Drug Administration

(Collins, 1978b) use the two-tailed t-test for litter size, mean liveborn per pregnant animal, and mean pup survivors postnatally. Fertility indices are analyzed by the two tailed Chi squared test. Viability, weaning, and survival indices may be transformed by using the Freeman-Tukey arc sine transformation for binomial proportions (Mosteller and Youtz, 1961) and are generally analyzed using the Dunnett's t-test. Data are sometimes reported on the basis of the proportion of the litters which are effected. The use of a nonparametric technique such as the Wilcoxon ranked sum test may provide a sufficiently powerful technique in the analysis of such data (Haseman and Soares, 1976). Indications of systemic toxicity, which are not complicated by litter effects, are analyzed using techniques such as Analysis of Variance and t-tests which are routinely used in the assessment of subchronic and chronic toxicity.

Different sets of statistical tests for reproduction studies may be used by other investigators and discussion of the appropriateness of each test is out of the scope of this evaluation procedure. When in doubt, it is suggested that the reviewer consult with HED statisticians.

I. Data Reporting/Final Report

Reporting requirements are listed in the Subdivision F Guidelines, §80-4 and §83-4, (USEPA, 1982) and will not be discussed in detail in this evaluation procedure. In general, the final report should contain tabulated data relative to parental body weight, parental organ weight, food consumption, parental clinical observations and mortality, reproductive results, pup survival data and indices, pup body weight, male/female sex ratio, parental and pup necropsy data, and parental and pup histopathologic findings (Tardiff and Rodricks, 1987; Dixon, 1989). The reviewer should be able to identify the dam and sire from the individual animal data and to associate all reported findings with individual litter data. All reported mean data should be carefully compared to submitted individual litter data for consistency across all generations and dose levels.

Submitted study reports should be signed and dated by the investigator(s); signed and dated quality assurance, Good Laboratory Practice (GLP), and "Flagging" Criteria statements should be included in the report (USEPA, 1986d). This, of course, is not necessary for reports published in the open literature, although a report from the open literature may have to be subjected to additional investigation prior to acceptance for regulatory purposes. If a study is not signed by the investigator(s) or if the histopathologic findings are not confirmed by a pathologist, it is assumed that the report is subject to change and does not represent the final position of the investigator(s). The reviewer

should note that the report is considered as a draft and does not yet fully meet regulatory requirements, i.e., is classified no higher than Core Supplementary Data. The final report should be carefully compared to the draft report when the final report is issued.

IV. STUDY INTERPRETATION

Most endpoints in the evaluation of a multigeneration reproduction study can be grouped as either effects observed in the offspring or in the parents. However, this division (and the separation of endpoints into reproductive and systemic toxicity) is somewhat artificial because of the interrelationship of many of the aspects of toxicity which are observed in the reproduction study. Effects on fertility, for example, may have effects on litter size which may in turn influence pup weight and development and subsequent viability. The reviewer should not attempt to categorize the forms of toxicity which are observed except under certain circumstances (see Section I.C.).

Comparisons of parameters such as reproductive indices can also be made between litters and between generations in a multigeneration study. However, in the design of a multigeneration reproduction study, the length of exposure between parental (P generation) and filial (F generation) is not equal. Treatment of the P generation commences with adult animals (at least 6 weeks of age), whereas filial (F1 and F2) generations are exposed continuously <u>in utero</u> from conception and throughout gestation. The F1 generation is also exposed and observed through puberty and into adulthood. Therefore, adverse reproductive effects observed in any particular generation are not necessarily indicators or predictors of similar adverse effects in another generation. Similarly, successive litters of the same generation (Fa and Fb) should not be treated as replicates (See Section IV.A.). Chemicals which bioaccumulate (have long half-lives for elimination and thus require prolonged exposure to reach steady state) may have a greater incidence or severity of effects observed in the second litter (F1b or F2b) as compared with the first litter (F1a and F2a) (Christian, 1986). Parity, age and body weight are also different at the time of the second mating.

A. Endpoints of Parental Toxicity

In the evaluation of a multigeneration reproduction study, it is important to assess whether an adequately high dose level has been used. The highest dose selected should produce some indication of maternal or adult toxicity, and all systemic endpoints should be considered. These can include increased incidence of mortality or other clinical signs, as well as significant changes in absolute

body weights, body weight gains, absolute and relative organ weights, feed and water consumption, clinical pathology, gross necropsy, histopathology, and/or cholinesterase activity data. No effects of toxicological significance should be observed in parental animals of the low dose group. If the data are available, the NOEL for systemic toxicity should be compared to NOELs in the chronic rat study to determine whether a greater sensitivity of the pregnant or lactating female is observed. See Section I.C. for further discussion.

1. Clinical Observations

Clinical observation data should include examination relative to the fur texture (i.e., matted, piloerection), skin (i.e, alopecia), eyes (i.e., mydriasis, miosis, nystagmus), mucous membranes (i.e., cyanosis), orifices (i.e., nasal discharge, vaginal bleeding, diarrhea), respiratory system (i.e., hyperpnea, dyspnea), autonomic and motor system (i.e., paralysis, paresis, fasciculation), behavioral changes (i.e., pica), and death. A careful evaluation of the reported clinical signs should alert the reviewer to treatment-related effects and allow a more accurate determination of the NOEL and LOEL.

Parental death or reduction in body weight gain relative to controls are obvious end-points of systemic toxicity and may result from many factors. Environmental factors, as discussed earlier, are known to influence the welfare of the test animals. Technical factors such as intubation errors, e.g., perforation of the esophagus/stomach or intertracheal administration on a gavage study, and mishandling of the animals may alter the outcome of the test results and lead to parental death and/or unnecessary additional stress superimposing pregnancy.

A necropsy should be conducted on all animals found dead during the course of the investigation. It is important to determine the cause of death if possible (accidental or compound-induced) to better understand the clinical toxicity induced in the dams by the test compound. The pregnancy status and the time of death (premating, mating, gestation, lactation, or rest period) should also be ascertained. Such information may be used to determine whether dams are sensitive to compound-induced toxicity during one or more of these periods. Maternal deaths, particularly at the high dose level, may influence the calculation of reproductive parameters and the interpretation of reproduction indices.

2. Parental Body Weights

Parental absolute body weights and body weight changes are recognized as sensitive indicators of systemic toxicity for most species. Unfortunately, they are nonspecific and also may result from anorexia induced by reduced palatability of the test diet, rather than actual toxicity. Reduced palatibility is suggested by decreased food consumption in the absence of other indications of toxicity. Confirmation, if necessary, may be obtained with paired feeding studies or with studies using an alternative route of administration. The body weight data should be reported on a weekly basis (at a minimum) for the growth period prior to mating as well as during gestation and lactation. Non-pregnant females should be excluded from the calculation of mean body weight and food consumption for the periods of gestation and lactation.

Body weight data are important for the determination of the NOEL. However, to be of greatest utility, all groups must have comparable body weights at the initiation of treatment. This criterium is managed by random selection of untreated animals for the P generation, but it is not always possible for the second generation in which some (treated) F1 pups may exhibit decreased body weight as a toxic effect or as a result of a potential 3-week difference in age. Decreased parental body weight gain is an effect which is usually seen in each generation, and it is recommended that trends in body weight be examined over the course of the entire study. However, it must be borne in mind that the P generation receives a shorter exposure to the test substance than the F1 generation and that their exposure does not encompass the period of perinatal development.

It has long been recognized that severe body weight loss can affect cyclicity in humans as well as in other mammalian species (Frisch, 1978; Merry and Holehan, 1979; Nelson and Felicio, 1984). However, in general, a modest reduction in body weight gain as a result of decreased appetite is not expected to have any significant effect on reproductive parameters (Zenick and Clegg, 1989). Recent feed restriction studies in experimental animals support this conclusion in rats but not necessarily in mice (Chapin et al., 1993a,b). A reduction in body weight may be due to a direct effect of the test material on the metabolism or digestive processes of an organism or may result from a decrease in food intake unrelated to the inherent toxicity of test material (decreased appetite or palatability). Therefore, body weight data should be assessed along with food consumption data for the calculation of food efficiency.

3. Food Consumption Data

If the test material is administered in the diet, the amount of food consumed is important in determining the exact amount of the test material received by the animal. In a multigeneration reproduction study, food consumption is measured on at least a weekly basis at different periods across all generations. The food consumption data are usually available for both parental males and females during the premating period and for parental females during the gestation and lactation periods.

A reduction in food intake may rarely be used to determine the NOEL; it is generally not observed in the absence of other indications of toxicity in the study. Food consumption varies from weaning to maturity, with younger animals consuming more food (on a kg/body weight basis). Unusually high food consumption data, which are actually due to spillage, are sometimes reported in toxicity studies. Very high consumption estimates in comparison to the Lehman tables (Table 3) suggest that spillage may have been a problem. The food consumption data during the lactation period are generally of questionable value because of significant amounts of spillage due to the pups playing in the food container and also since the dam and her unweaned litter eat from the same container late in lactation. It should be also noted that a decrease in food consumption or body weight may be due to other factors unrelated to the test chemical. Nontreatment-related diarrhea, disease, and decreased water consumption may be cited as some of the possible causes.

Food consumption may be expressed as g/animal/day or g/kg body weight/day. The latter is preferred since the former excludes differences in body weights which may occur among the groups. As indicated earlier, body weight and food consumption data are best evaluated together since these parameters are interdependent. Evaluation of the body weight data along with the food consumption data may provide the reviewer with information relative to either a change in appetite or palatability or a change in the food efficiency. The latter is a measure of the efficiency of the food utilization, which can be calculated as follows:

$$\frac{\text{grams body weight change per unit time}}{\text{grams food consumption per unit time}} \times 100$$

This calculation gives the percentage efficiency with which the animal converts food for maintenance. Low efficiency compared with the controls indicates toxicity in the consuming animals, i.e., if the mean food efficiency value is similar between the treated and control groups, then anorexia may not be the main factor in depressing the body weight.

4. Compound Consumption Calculations for Dietary Studies

Although the EPA Guidelines do not provide guidance regarding compound consumption data on reproduction studies, these data are often provided in study reports for dietary studies. Individual and mean daily compound consumption values (mg compound/kg body weight/day) are calculated from the body weight and food consumption data, utilizing the nominal concentration of test material in the diet at each dose level. Generally the data are collected weekly, and since the animals consume feed and gain body weight at different rates during their life spans, the compound consumption of each dose group will change over time. For a multigeneration study, mean daily compound consumption values are generally presented by dose, sex, and generation over several defined phases of the study, e.g., the premating phase (approximately Weeks 0-10 of study for both males and females) and gestation and lactation (for pregnant females only).

The apparent variability of the mean compound consumption data within each study (by sex, by study phase, by generation, and over time during growth) requires the reviewer to apply discretion in selecting the most appropriate value as representative of study test substance consumption for each dose level.

It is not appropriate to use maternal gestation or lactation compound consumption data as representative of the entire study, since the dams (and eventually the weanling litters) consume unusually large amounts of feed while in these reproductive phases. Therefore, the data from the premating period, in which the animals (both male and female) are on study from the time they are weanlings until they are sexually mature adults, are the optimal data set. Generally, overall (i.e., Weeks 0-10 or Weeks 0-12) values are calculated and presented in the report; these are usually calculated from individual animal overall values. If only individual weekly values are reported, the reviewer can calculate an average of the weekly mean values to estimate overall mean compound consumption. Differences in the body weight and food consumption data between sexes require that the reviewer summarize compound consumption data separately by sex.

The following principles can be used to determine which generation should be used to represent average compound consumption for the study:

a. If there are obvious differences in toxic effects noted between generations, and the NOEL and/or LOEL are based upon the data from only one generation, the calculated compound consumption values from the most affected generation should be used to define the doses.

b. If there are no obvious differences in toxic effects between generations, the most conservative set of values should be used. In other words, if the compound consumption values for one generation are clearly lower than those of the other, the lower values should be used to represent the study.

If no compound consumption values are reported, the species-specific nominal conversion factors recommended by Lehman (USFDA, 1959) should be used to convert dose level values from ppm to mg/kg/day (see Section III. E., Table 3).

5. Water Consumption Data

Water consumption data are rarely available (and are not specified by guideline), unless the test material is administered via drinking water, since most test facilities utilize an automated water supply device which does not record individual consumption. In the event that water consumption data are recorded, an increase may suggest the possibility of renal toxicity. However, this finding should be corroborated by necropsy observations, histopathologic changes, or kidney weight changes to provide a conclusive endpoint of systemic toxicity. In rodents, a decrease in water consumption may lead to dehydration, decreased food intake, and body weight reduction, which eventually may lead to a sequelae of adverse reproductive outcomes.

6. Necropsy Observations

The 1982 FIFRA Guidelines state that "a complete gross examination should be done on all animals, including those which died during the experiment or were killed in moribund conditions" and "special attention should be directed to the organs of the reproductive system". Uterine implantation sites can be counted to provide the means of estimating postimplantation loss by comparison to the pup count at birth. Postimplantation loss can be estimated accurately for dams killed after production of one litter only; the identification of corpora lutea and implants related only to the second pregnancy is more difficult for the second litter.

Necropsy data should be tabulated per group and generation. From these data it is essential to attempt to ascertain whether technical errors (e.g., gavaging errors), diseases, or the test material toxicity itself, are responsible for the observed mortalities. For example, reddening of the trachea, congested lungs, and fluid accumulation in the lungs are highly suggestive of gavage errors and/or diseases. Further, not only the cause of death (accidental or compound-related)

must be determined from the necropsy data but the pregnancy status of these animals must also be ascertained.

7. Organ Weight Data

Organ weights for the reproductive and nonreproductive organs are not required by the 1982 EPA or 1983 OECD test guidelines; as a result, they are often not available. However, organ weights are a useful component of the macroscopic examination, since they provide the first signs of dystrophic or dysplastic changes. Measurements of organ weights are easily incorporated into study protocols. In the absence of organ weight measurements from reproduction studies, these data from other relevant and available toxicology studies may provide valuable information.

If included in the study design, organ weights should be expressed on both an absolute and relative basis. The relative organ weight takes into account the difference in body size (terminal body weight), since organ size generally increases with body size. However, the increase in organ weight is not directly proportional to body weight but instead is more closely related to surface area. The expression of the relative organ weight may not be biologically accurate in the presence of significant differences in terminal body weight among the groups. The interpretation of relative organ weights should be limited to groups of animals with comparable terminal body weights. In fact, the body weight reduction observed in the treated groups is frequently due to a reduction in fat deposition and not necessarily due to a depression in the development of lean body mass. Stevens and Gallo (1989) suggested that when significant treatment-related differences in a study are detected in many organs relative to body weight, organ/brain weight ratios should be analyzed, since subsequent to development the weight of the brain remains quite stable in adult animals.

Reproductive organ weights are not affected by slight body weight decreases (Chapin et al., 1993a,b). Data on nonreproductive organ weight (e.g., liver, kidney, adrenal glands, brain, spleen, and other known or potential target organs) not only provide the reviewer with information relative to the target organ toxicity of the agent being tested but also is useful for the determination of the adequacy of dosing. Such information should be used in conjunction with organ weight data from subchronic and chronic studies in the same species to assess systemic toxicity.

Although not mentioned in the 1982 Guidelines, data on reproductive organ weights should be collected for both male and female animals. Reproductive

organs of interest are the ovary and uterus for females and the testis and the accessory sex glands (epididymis, prostate, and seminal vesicle) for males. In addition, adrenal gland weight may be a sensitive indicator of some hormonally-mediated effects.

Significant alterations in ovarian weight may suggest a female reproductive toxic effect. However, some investigators have noted that ovarian weight varies with the stages of the reproductive cycle and the number of corpora lutea present at sacrifice. Similarly, alterations in uterine weight should be evaluated carefully since the weight of this organ is under significant hormonal influence. Evaluation of uterine weight data (and histopathology) in conjunction with data on the stage of the estrous cycle at sacrifice allows more meaningful interpretation; however, these data are not usually provided.

Testicular weight data collected during chronic testing may provide an indicator of male reproductive toxicity, particularly those weights recorded at the time of the one year interim sacrifice. Testicular weights of old rats at final sacrifice in chronic studies are often complicated by a high background rate of atrophy of tubule and/or Leydig cell hyperplasia (near 100% for Fischer rats). Although testicular weights do not exhibit a great variability between animals (Blazek et al., 1985), they are not a particularly sensitive indicator of testicular toxicity (Foote, 1986). However, sometimes they can be treatment-related (e.g., increased testes weight resulting from blocked efferent ducts following treatment with carbendazim; Gray et al., 1990). Among the male reproductive organs, the testis shows the least weight variation among normal members of a given species (Schwetz et al., 1980; Blazak et al., 1985). In the rat, testes grow at the same rate as the body. In addition, with senescence, there is a decrease in gonadal weight. The epididymis, prostate, and seminal vesicle weights may provide valuable information if care is taken in their removal and dissection. The accessory sex glands (seminal vesicles and prostate) can be weighed with or without fluids and are very sensitive to decreased androgen stimulation (as are the testes).

Pituitary and accessory sex organ weight should also receive attention as it may provide information regarding the target site or most sensitive endpoint for reproductive toxicity. Some physiologic functions of the pituitary gland are unrelated to reproduction. Hence, alterations in pituitary weights may be considered as adverse effects but do not necessarily reflect adverse reproductive effects (for either males or females). In general, changes in the pituitary are not considered to be more sensitive endpoints than those observed in other reproductive organs; therefore, lack of pituitary weight data does not usually

compromise the study results.

8. Histopathological Findings

Histopathological information is required and should be provided in the study report for all high-dose and control parental (P) and filial (F1) animals selected for mating and in all organs showing macroscopic changes (USEPA, 1982). If treatment-related histological findings are observed at any site, examination of those tissues from the mid- and low-dose levels is required. A list of tissues required for histopathologic examination is given in §83-4 of the FIFRA Guidelines (USEPA, 1982).

Grading of the lesions is often necessary to assess findings, particularly with commonly occurring histological changes, although this information is often not provided. As for any other toxicological findings, a dose-response relationship in the frequency and/or severity of the findings assists in determining whether the effects are treatment-related. A compound-related effect may also be established if significant changes are observed only in the highest dosage level group.

Histopathological examination of the testes may include an evaluation of the spermatogenic process through examination of the stages of the seminiferous epithelium (Russell et al., 1990). This amount of detail is not necessary if caudal epididymal sperm are evaluated for concentration, morphology, and motility; however, these data are not required and are usually not provided. Serial histopathological examination of the ovaries to quantify oocytes may be performed and can provide additional information on oocyte depletion either in the presence or absence of other hormonally-mediated effects in the females. It is a laborious process but no other method is currently available to detect oocyte loss.

Histopathological changes in the gonads, particularly in males, may not be correlated with alterations in fertility since there is often a lack of sensitivity of the rodent to decrements in sperm count and standard measurements of fertility do not identify these effects. Thus, histopathological changes may be observed a dose levels at and below which a decrease in fertility is evidenced. Definitive conclusions on possible adverse effect of the reproductive organs can only be made with functional tests which analyze the response of the endocrine/reproductive system. Possible use of functional tests may be discussed in future Guidelines. In a reproduction study, morphological changes in both the endocrine and reproductive systems must be carefully monitored. Interpretation

of histological changes at these sites can be facilitated by recent publications such as that of Russell et al., (1990). Any histopathological change in reproductive tissue should be regarded as of concern.

9. Cholinesterase Determinations

After dosing with known cholinesterase-inhibiting pesticides, cholinesterase levels are often measured in the plasma, red blood cells, and brain of parental animals. Cholinesterase measurements are rarely available for neonatal animals. If cholinesterase data are available and demonstrate pronounced levels of toxicity which are both biologically and statistically relevant in parental animals, these data may be used as a basis for determining whether the dose levels utilized were adequate for the investigation of reproductive toxicity. At lower dose levels, such measurements may also be considered in the establishment of a NOEL. Measurements of cholinesterase in neonates or fetuses may indicate a preferential sensitivity of the young, as has been suggested by studies performed with aldicarb (Cambon et al., 1979). It has been recommended that the measurement of brain cholinesterase in pups be routinely conducted for pesticides with anticholinesterase activity (FAO, 1990).

10. Reproduction Data

Reproductive indices and other endpoints are assessed for parental animals from mating to parturition/sacrifice and may provide information relative to the effects of an agent on libido, germ cell alterations, gametogenesis, fertilization, estrous cyclicity, implantation, embryonic growth, survival, and histopathology changes. Caution should be exercised when reviewing reproductive indices to ensure that the same definitions for the indices are being used by the registrant and the toxicology reviewer. This will allow a more consistent interpretation of possible reproductive changes.

a. Male Mating Index

The male mating index is defined as:

$$\frac{\text{Number of males for which mating was confirmed}}{\text{Number of males used for mating}} \times 100$$

Although this index is often not reported, its calculation is feasible from the individual litter data submitted. It can be calculated for each generation and at each mating (Fa and Fb) of subsequent generations. It provides information relative to the number of sexually active males (ability to mate) in

those studies where treated males are mated to untreated females in order to further assess fertility problems. However, since both sexes are usually treated in multigeneration studies, this index is not a specific measurement of male reproductive toxicity and can only suggest effects on male fertility.

Mating is confirmed on the basis of the presence of vaginal plugs, plugs in the pan beneath the animals, or sperm in vaginal lavage. To truly ascertain the male mating index, pair mating (1 male: 1 female) is preferred. A decrease in the male mating index may be due to many factors, which include but are not restricted to absence of libido, hormonal imbalance, estrous cycle disruptions in the female, or impotence of either sex. The etiology for a decrease in male mating index may be due to alterations in either the sensory, motor, hormonal or autonomic system. Regardless of the cause, evidence of a dose-response with or without persistent effect throughout subsequent generations is indicative of treatment-related effects in the male. A reduction in male mating index should be carefully evaluated along with possible histopathologic changes in the male reproductive organs.

Mating indices are useful because they can yield additional information about the integrated function of the neuroendocrine-gonadal axis. Evidence for an adverse mating behavioral effect in animals is considered suggestive of a potential for an adverse effect on human reproductive function.

b. Male Fertility Index

The male fertility index gives an indication of the outcome of mating and is calculated as follows:

$$\frac{\text{Number of males impregnating at least one female}}{\text{Number of males used for mating}} \times 100$$

In a rodent study, mating is normally determined by the presence of a plug (copulatory or vaginal) and/or presence of sperm in the vaginal smear. However, neither of these indicators necessarily ensures that pregnancy will ensue. In fact, a copulatory plug is only a product of secretions of the vesticular and coagulating glands of the male and does not necessarily indicate the deposition of sperm. The presence of sperm in the vaginal lavage does not imply that fertilization and/or implantation will occur. Further, the male rodent must provide an adequate number of intromissions of sufficient force and an adequate number of ejaculations for sperm transport and for the female to respond with sufficient uterine contractions and released progesterone for the initiation of pregnancy (Adler et al., 1970; Chester and Zucker, 1970).

c. Copulatory Index

$$\frac{\text{Number of estrous cycles with copulation}}{\text{Number of estrous cycles required for pregnancy}} \times 100$$

The male fertility index is also rarely reported by the study author(s) but can be calculated from the supporting data. The male fertility index provides information relative to the number of proven fertile males, but in studies in which both sexes are treated, it is not specific for male reproductive toxicity.

The number of estrous cycles or the length of time required from initial pairing to copulation must be determined for each female from the supporting data. These data should contain information relative to male and female pairing, date of mating, number of copulations observed, and/or the number of estrous cycles required. Rats and mice usually allow mating to occur only at estrus (which lasts from 12-14 hours in 5 day cycles). Although rodents are spontaneous ovulators, ovulation does not necessarily accompany every estrous cycle (Jacoby and Fox, 1984). Normally, rodents require only one estrous cycle for mating to occur, and an increase in the number of estrous cycles that a female would require is suggestive of subfertility. Data on the estrous cycle may provide some indication of subtle changes in endocrine status, since the cyclic changes in vaginal cytology reflect the changes in endocrine milieu (Schwartz et al., 1977).

d. Female Fertility Index

The female fertility index is defined as:

$$\frac{\text{Number of females conceiving}}{\text{Number of females cohabited with males}} \times 100$$

This index provides a general measure of female fertility, one which is dependant upon the definition of "females conceiving" and may be constrained by laboratory standard operating procedures or study-specific protocol requirements.

The calculation of female fertility may include any of the following definitions of the number of females conceiving. Clearly, there is a resulting potential for distinct differences in data interpretation.

1. The number of females conceiving = the number of females with positive evidence of copulation. This variation of the calculation uses the (generally erroneous) assumption that all females that mate

are pregnant. Calculated on this basis, the index could more accurately be described as a female mating index; yet this may also be incorrect because the calculation does not include females for which evidence of copulation was not observed.

2. The number of females conceiving = the number of females that delivered litters. This calculation assumes that fertility is equivalent to the production of a term litter. It is more closely representative of the (parental) birth index, which is defined as:

$$\frac{\text{Number of females with live born}}{\text{Number of females cohabited with males}} \times 100$$

although full-term litters which consist entirely of dead pups are discounted from the offspring live birth index. The parental birth index should not be confused with the offspring birth index, which is discussed in Section IV.B.1.a.

3. The number of females conceiving = the number of females that delivered (or aborted) litters. This version of the formula does not include the number of females which mated but which did not become pregnant. For that reason, it also is more an indication of pregnancy than of fertility.

4. The number of females conceiving = the number of females that delivered (or aborted) litters **plus** the number of females which had totally resorbed litters. In order to determine whether a female was pregnant but did not deliver a litter due to the gestational death and resorption of the entire litter, the uteri of apparently nonpregnant females are examined at study necropsy, often including a staining procedure (using ammonium sulfate) in order to identify nidations. Although this formula may result in a more accurate index than the calculation described in definition #2, above, it is still a measurement of pregnancy rather than fertility.

5. The number of females conceiving = the number of females with positive evidence of copulation **plus** the number of females that delivered (or aborted) litters, even though no evidence of copulation was observed (double counting is excluded). This formula includes mating events as well as pregnancy outcome, and it negates the effect of technical error in determining positive signs of copulation as long as clear signs of the presence of a litter are

noted. However, it does not account for litters which are totally resorbed (and for which there was no evidence of copulation).

6. The number of females conceiving = the number of females with positive evidence of copulation **plus** the number of females that delivered (or aborted) litters **plus** the number of females which had totally resorbed litters, even though no evidence of copulation was observed (double counting is excluded). As stated previously, determinations of total resorption of litters are made by uterine examination at necropsy. This calculation includes measurements of both mating and pregnancy (accounting for nearly all possible outcomes), and is the best indicator of female fertility.

The FIFRA Guidelines (USEPA, 1982) state that females unmated after 21 days (first mating trial) may be paired with proven fertile males of the same treatment group (second mating trial). The use of proven males for a second mating is desirable because such use minimizes interpretation errors that can arise if inexperienced males are used. Rodents not achieving pregnancy after a second mating trial are considered to be infertile, since they have had repeated opportunities to mate with a fertile male. Female fertility indices from studies in which remating of presumed infertile females was conducted should not be compared to those same indices from studies (or generations of the same study) which did not include a second mating.

e. Fecundity Index

This index reflects the percentage of matings resulting in confirmed pregnancies and is expressed as:

$$\frac{\text{Number of pregnancies}}{\text{Number of copulations}} \times 100$$

The fecundity index reflects the total number of dams that have achieved pregnancy, including those that deliver at term, abort, or fully resorb their litters. The main difference between this index and the female fertility index is that the fertility index measures ability of the females to produce a litter when exposed to (paired with) the males whereas the fecundity index measures the ability of the females to become pregnant after mating to (copulation with) the males. As mentioned previously, copulation (presence of plug and/or sperm in vaginal lavage) does not necessarily ensure that fertilization and implantation will occur.

As previously noted, one of the limitations of the multigeneration reproductive study is the inability to distinguish whether the reproductive adverse effects, if present, are primarily due to the male or female animals. A careful comparison of the male fertility index and female fecundity index and accompanying histopathology or other data, may provide the reviewer with some information.

f. Gestation Index

The gestation index is a measure of the efficiency of pregnancy resulting in at least one live offspring. The gestation index is defined as:

$$\frac{\text{Number of females with live born}}{\text{Number of females with confirmed pregnancy}} \times 100$$

This index is of limited sensitivity because litters with one or more live pups are treated as of equal biological significance. Therefore, although it is a measure of the number of litters with live offspring, it is an incomplete measure of fetal mortality unless the whole litter succumbs (Collins, 1978b). To clarify the effects of an agent on pup viability at birth, it is recommended that the reviewer examine the live birth index (see Section IV.B.1.).

g. Mean Time-to-Mating

At each breeding, time-to-mating values can be determined for individual breeding pairs and a mean calculated by dose group. The time-to-mating is defined as the number of days from initiation of cohabitation until positive evidence of copulation is observed. The time-to-mating for normal rodent pairs could vary by 3 or 4 days, depending upon the stage of the estrous cycle in the female at the time of pairing, so the reviewer should be aware that some variation in the individual data is normal and acceptable. Treatment-related increases in time-to-mating values could indicate reproductive abnormalities such as abnormal estrous cyclicity in the female or impaired sexual behavior in one or both partners. Increased time-to-mating will not necessarily result in negative changes to other fertility indicators such as the mating, fertility, or gestation indices.

h. Mean Gestation Length

The length of gestation is the interval from positive evidence of copulation to parturition. It can be calculated from the individual breeding records of the animals and is summarized as a mean value. Significant decreases

in gestation length may result in decreased offspring birth weight and survival. Significant increases may result in dystocia (difficult labor and delivery), and ultimately to the death or physical impairment of the dam and/or offspring. In addition, if lengthened gestation is not complicated by dystocia, the offspring may be larger and heavier at birth, which could, without the application of more sophisticated statistical analysis techniques, mask treatment-related body weight gains of the offspring later during lactation.

i. Sperm Evaluation for Male-Mediated Effects

Sperm measurements, although not yet required by FIFRA guidelines (1982), may provide valuable information in cases where reproductive toxicity is male-mediated. Although sperm measurements are most readily obtained from species larger than rodents, data concerning sperm production and characteristics are increasing being gathered from subchronic, chronic, and reproduction studies in rats, and it is important that this information, if available, be carefully considered in the evaluation of reproductive toxicity. See Russell et al.(1990) for a discussion of the interpretation of sperm measures.

It is preferable to use caudal epididymal sperm (collected at necropsy) for evaluation. Samples taken by electroejaculation are not acceptable (Zenick and Clegg, 1989). The latter technique may result in sperm measurements that are not reflective of the sperm that may be delivered through normal copulation. If a treatment-related effect is observed in the cauda epididymal sperm concentration, the spermatid count may be evaluated. The variances of cauda sperm versus spermatid (tid) counts should be compared; tid count variance tends to be higher and is therefore not likely to be as sensitive.

The production of sperm may have to be severely reduced, e.g., by 80 to 90% in some strains of rodents, before fertility is effected in the multigeneration reproduction study (Robaire et al., 1984; Working, 1988). Human fertility, on the other hand, may be effected by a small decrement in sperm production at least partially because many men, especially those over the age of 30, have daily production rates of normal sperm which are inadequate or barely adequate to ensure fertility (Amman, 1981). Direct measurements of sperm number and quality in rodents are more sensitive and specific indicators of toxicity than measures of fertility. An additional advantage of performing sperm measurements is that these data can also be obtained from humans in situations where animal studies suggest a hazard, enhancing the ability to confirm the extrapolation of test results to humans.

Two reviews of sperm evaluations have been published by the EPA Gene-Tox

Program (Wyrobeck, 1983a,b). Other reviews of the use of sperm measurements in rodents are included in articles by Amann (1986), Blazek et al. (1985), Zenick and Clegg (1989) and Russell et al. (1990). These references should be consulted for detailed guidance on the interpretation of sperm measurements and for a compilation of reference values. Reproductive toxicologists in the Health Effects Division (HED), the Health Effects Research Laboratory (ORD/HERL) and the Office of Health and Exposure Assessment (ORD/OHEA) can also provide guidance concerning the interpretation of sperm measurements. The most frequently collected sperm measurement includes sperm count (as an indicator of sperm production), sperm morphology (which generally includes only the evaluation of head shape and is thought to be related to the ability of a sperm to reach and fertilize the oocyte), and sperm motility, expressed as percent progressively motile sperm (which is influenced by a large number of variables and is an indicator of both sperm maturation and ability to survive and fertilize). The use of video technology in sperm assessment is preferable, to provide a permanent record of the sample collected and allow more accurate measurement of the data. General species differences for sperm parameters are shown in Table 7.

j. Cyclicity Data in the Evaluation of Female-Mediated Effects

The evaluation of female cyclicity data can be helpful in identifying female-mediated effects on fertility. These data may be included in a study report even though they are not required by the 1982 FIFRA guidelines. Samples of vaginal epithelium are collected daily by lavage and examined microscopically to determine the stage of estrous cycle (see Table 6 below).

Table 6. Estrous Cycle in Rats (4-5 day cycle)

Stage	Duration (hrs)	Microscopic Characteristics of Vaginal Lavage
Proestrus	≈12	Round-oval cells
Estrus	≈12	Maximal fluid distension, cornified cells
Metestrus	≈21	Cornified cells and leukocytes
Diestrus	≈57-70 (2.4-2.9 days)	Leukocytes only

Although it is possible to collect cyclicity data daily throughout the study, evaluation of specific points in the reproductive life of the study animals should yield sufficient data for evaluation of effects. The onset of female cyclicity can be determined by daily smears of puberty-age females (after vaginal opening); this milestone in reproductive development can also be evaluated by examination of the age at which vaginal opening occurs, although normal cyclicity is not necessarily indicated by this event. Evaluation of vaginal smears during the mating period can sometimes offer explanations for a

Table 7. Species Differences in Spermatogenesis, Daily Sperm Production, and Epididymal Transit Time[a]

Parameter by species	Mouse	Hamster	Rat	Rabbit	Dog (beagle)	Monkey (Rhesus)	Man
Duration of spermatogenesis (days)	34-35	35-36	48	48-51	62	70	72-74
Duration of cycle of seminiferous epithelium (days)	8.9	8.7	12.9	10.7	13.6	9.5	16.0
Life span (days)							
B spermatogonia	1.5	1.6	2.0	1.3	4.0	2.9	6.3
Leptotene	2.0	0.8	1.7	2.2	3.8	2.1	3.8
Pachytene spermatocytes	8.0	8.1	11.9	10.7	12.4	9.5	12.6
Golgi spermatids	1.7	2.3	2.9	2.1	6.9	1.8	7.9
Cap spermatids	3.6	3.5	5.0	5.2	3.0	3.7	1.6
Testicular wt. (total)	0.2	3.0	3.7	6.4	12.0	49.0	34.0
Daily sperm production (millions)							
Per gm of testis	28	24	24	25	20	23	4.4
Per male	5.6	72	89	160	240	1127	150
Sperm reserves in caudae epididymides at sexual rest (millions)	49	1020	440	1600	2100	5700	420
Epididymal transit time at rest (days)		14.8	8.1	12.7	11.3	10.5	5.5-12

a Data derived primarily from Amann (1986) as cited in Biologic Markers in Reproductive Toxicology (NRC, 1989).

female-mediated fertility reduction in the mated pair. For females that show no evidence of copulation while paired with a male, the duration of the mating period can be defined or regulated by the number of full or partial estrous cycles that pass. Knowledge of the stage of the estrous cycle at the time of necrosy can aid in the interpretation of female necropsy and organ weight data, particularly for uterine observations, since the weight and histology of the uterus is hormone-dependent. Taking this procedure one step further, some protocols may even require that all females be sacrificed at the same stage of of the estrous cycle in order to facilitate interpretation of the data.

The reviewer should be aware that care must be taken in the laboratory to avoid stimulation of the cervix with the pipette during the vaginal lavage procedure, or pseudopregnancy (as evidenced by prolonged diestrous) may be inadvertantly induced, and fertility may be artifically diminished. This lack of good laboratory technique could potentially compromise the results of an otherwise adequate study.

B. Endpoints of Offspring (Filial) Toxicity

Offspring are assessed from parturition, and through weaning and/or sacrifice and may provide information relative to the effects of an agent on neonatal growth and survival, lactation, postnatal growth (weights, crown-rump length), developmental milestones and histopathology.

1. Litter Size

Mean litter size is a reproductive parameter that should always be considered in the evaluation of data. The mean litter size for each treatment group should be calculated for each observation interval, for day 0 it is defined as:

$$\frac{\text{Total number of pups delivered (live plus stillborn)}}{\text{Number of dams that delivered}}$$

At lactation intervals subsequent to day 0, mean litter size should be defined as:

$$\frac{\text{Total number of live pups}}{\text{Number of dams}}$$

In the calculation of the day 0 mean litter size, pregnant females that die or abort are excluded. It is more appropriate to use the total number of pups born (stillborn plus live pups) in the calculation of this mean instead of only the total number of live pups as reported by many investigators since the stillborn pups were viable entities of the litter throughout gestation. This provides a valid number to compare with other parameters such as the mean litter weight or maternal body weight gain during gestation.

The day 0 mean litter size value represents the mean number of live pups per litter based on the total number of females presumed pregnant. Determination of the pregnancy status of all dams is, therefore, crucial in calculating this value, since the number of females presumed pregnant should include those which have had litters fully resorbed. A decrease in the live litter size may be indicative of a reproductive effect, resulting from a decrease in numbers of oocytes ovulated, an increase in fertilization failure, an increase in the number of pre- or postimplantation losses, or an increase in the incidence of pregnant females which fail to deliver or have viable offspring. An increase in pre- and postimplantation loss may be, but is not necessarily, the result of a genotoxic effect of the chemical. However, it is usually impossible to distinguish between

failure of fertilization and death of the zygote.

Mean litter size is often affected by cannibalism. Evidence of cannibalism is generally included within reported offspring clinical observations, indicating the number of partially cannibalized (dead) pups found per litter at each observation period. However, cannibalized pups are often completely consumed by the dam, particularly in early gestation when the pups are rather small, leaving no evidence to indicate the fate of the missing offspring. Other than examination of the uterus for empty implantation sites immediately after birth and comparison to total litter size, there is no way to precisely determine the number of pups cannibalized at birth on a typical two-generation reproduction study; of course, this procedure is not possible for a guideline study since the dams are required to raise the litters to weaning. Close monitoring of delivering dams by the laboratory technical staff and immediate removal of dead pups from the nesting cage may prevent loss of some information, yet the reviewer often has no way of determining whether or not these procedures were instituted. The reviewer will, however, be able to identify complete postnatal cannibalization of pups as discrepancies between litter counts at birth and on subsequent observation periods during lactation. These numerical reductions in litter size will not be accounted for by reported pup deaths or losses through standardization procedures on lactation day 4. By critical examination of individual and summary litter data, the reviewer should be able to identify increased incidences of cannibalism and determine if there appears to be a relationship to treatment.

A decrease in litter size may be indicative of possible adverse effects on either parental animal. In the males, this may include adverse changes in sperm quality and/or quantity. In females, adverse effects may have occurred during oogenesis, ovulation, fertilization, transport or implantation, and development of supportive organ systems, e.g., placentation.

A paternally-mediated effect on litter size can usually not be ruled out using only the information which is routinely available in the reproductive toxicity study. If available, information on the numbers of oocytes ovulated (recently formed corpora lutea) and implantation sites is necessary to evaluate the extent of pre- and early postimplantation losses. These data may provide additional information that can be used in the overall evaluation of female reproductive effects but can only be collected from animals which are sacrificed in late gestation.

2. Survival

a. Fertilization Efficiency (Pre-Implantation Loss), Implantation Efficiency (Post-Implantation Loss), and Fetal Gestation Viability Index

From the necropsy data of parental animals at weaning of the last litter of each generation, at least three indices can be calculated and are useful to study the effects of an agent on litter size. These are the fertilization efficiency, implantation efficiency, and fetal gestation viability indices.

The fertilization efficiency for each dam is defined as:

$$\frac{\text{Total number of implantations}}{\text{Total number of corpora lutea}} \times 100$$

and the implantation efficiency for each dam is a measure of:

$$\frac{\text{Total number of pups born (stillborn and live)}}{\text{Total number of implantations}} \times 100$$

Whereas the fetal gestation viability index is calculated as:

$$\frac{\text{Total number of live born pups}}{\text{Total number of implantations}} \times 100$$

Decreases in the fertilization efficiency and implantation efficiency indices represent, respectively, preimplantation loss and postimplantation loss. An evaluation of the above indices provides information concerning effects on fertilization, implantation, and early and late deaths of the zygotes.

b. Live Birth Index

This index is defined as:

$$\frac{\text{Number of pups born alive}}{\text{Number of pups born (total)}} \times 100$$

At birth, all pups should be examined for external anomalies as well as for viability. The number of viable, stillborn, and cannibalized members of each litter should be recorded (National Academy of Sciences, 1977). Distinction between stillborn (dead *in utero*) and pups which died shortly after birth (live born) is not always feasible since, from a practical point of view, the viability status of all offspring at birth cannot always be performed immediately after

parturition. However, this distinction can still be made in the laboratory by removing the lungs of dead pups and immersing them in water (the lungs of liveborn pups will float due to the presence of inhaled air). This distinction is of importance in the case of whole litter death; assuming that all members of a litter are stillborn may modify the outcome of the parental gestation index, which is defined as the percentage of litters with one or more live pups (Section IV.A.10.f.). A decrease in live birth index, therefore, reflects treatment-related effects manifested primarily during the advanced stages of pregnancy and resulting in stillborn.

Cannibalization is another problem which may arise and obscure the results obtained. Cannibalization is a behavioral change in the dams, which may be associated with general types of stress and is also a response by the dams to the delivery of malformed offspring. Inadequate food and/or water supply, elevated ambient temperatures, sudden changes in environmental conditions, lack of bedding, or poor animal handling techniques may be considered as part of the spectrum of stress-related changes (Harkness and Wagner, 1977). Cannibalization tends to be more prevalent among animals whose litters have developmental defects or are incompatible with life. Cannibalization may change the "number of pups born" used in the calculation of the live birth index (see Section IV.B.5. for additional discussion on cannibalism).

c. Viability Indices (Days 0-4 Survival)

The ability of the pups to survive is a primary focus in a reproduction study. Offspring viability can be impaired by:

i. Developmental effects of the young (abnormal and/or inadequate organ development) as a result of <u>in utero</u> exposure
ii. Varied effects of maternal toxicity:
    1. Maternal neglect (behavioral change)
    2. Inadequate milk production (endocrine change)
iii. Postnatal toxicity due to the presence of the agent in the milk during the lactation period.

Offspring viability indices are usually measured on postnatal days 4, 7, 14, and 21, and each index represents a specific period of the animal life. The reviewer should note that the term "viability index" can be employed under two definitions. Fitzhugh (1968) refers to it as the percentage of all young born that are able to survive 4 days, but other investigators use this term to indicate the survival ability of the pups to other time points in postnatal life.

Therefore, the meaning of "viability index" should be ascertained in connection with each report in which it is used.

For consistency, it is suggested that the viability index be defined as a measure of 4-day survival:

$$\frac{\text{Number of pups alive on lactation day 4 (prestandardization)}}{\text{Number of pups born alive}} \times 100$$

Offspring deaths occurring at any time prior to day 4 may contribute to a decrease in this index. A careful examination of the individual litter data may provide the reviewer with some indication of the etiology of the death. Early deaths in the offspring (i.e., postnatal days 0-1) are suggestive of functional defects (pulmonary, cardiovascular, or renal) rather than nutritional deficiency. Dead pups should be preserved and studied for possible defects and cause of death.

Data from some laboratories may indicate the presence or absence of a "milk spot" at the examination of moribund or dead pups. This is generally defined as a milk-filled stomach, which is externally visible through the thin, translucent skin of the pups. In addition, necropsy data may remark upon the presence of milk in the stomach. This information can be useful in attempts to distinguish between deaths resulting from lack of maternal care or pup nursing ability and deaths resulting from toxicity to test substance in the milk or other causes.

On postnatal day 4, the litter size is sometimes standardized. If the litters are standardized on day 4, the number of pups alive on lactation day 4 prior to standardization is used to calculate the viability index. It should be noted that although the 1982 FIFRA Guidelines indicate that all litters should be adjusted to 8, some other investigators prefer standardization to 10 to correspond with the number of mammary glands in rodents (3 pairs in the cervicothoracic region and 2 pairs in the inguinoabdominal region). In addition, some investigators (Palmer, 1986) believe that standardization is unnecessary and that it reduces study sensitivity (see Section IV.D.); the OECD two-generation reproductive toxicity test guideline No. 416 (1983) does not recommend standardization. However, lack of standardization does not necessarily imply that the study should be classified as Core Supplementary Data (see Section VI.A.5.).

d. Lactation Index (Days 4-21 Survival)

The lactation index is the viability index at postnatal day 21:

$$\frac{\text{Number of pups alive on day 21}}{\text{Number of pups alive on day 4}} \times 100$$

If litter standardization is performed, then the denominator is the number of pups kept after standardization on day 4. A decrease in the lactation index may likely result from in utero induced developmental effects, nutritional deficiency (perhaps resulting from an endocrine change in the dams), toxicity of the chemical (excreted in the milk), or maternal neglect. Regardless of the exact cause of death, an impairment of either the viability index or lactation index is considered to be a reproductive effect.

e. Weaning and Preweaning Indices (0-21 Days Survival)

In many nonstandardized studies the weaning index is reported instead of the lactation index. The reviewer should be aware that when the weaning index is calculated for litters that are standardized, the indices are invalid.. The weaning index is a measure of:

$$\frac{\text{Number of pups alive day 21}}{\text{Number of pups born alive}} \times 100$$

The weaning index gives an overall offspring survival data from birth to weaning. Another approach to the same issue is the determination of the overall offspring mortality in each litter, in which the preweaning index is calculated as:

$$\frac{\text{Number of viable pups at birth - number of viable pups at day 21}}{\text{Number of viable pups at birth}} \times 100$$

3. Sex Ratio

The sex of the offspring on a two-generation study is generally determined at birth, although the 1982 FIFRA guideline does not specifically discuss collection of this information. The data are presented in the report as either a sex ratio:

Number of male offspring : Number of female offspring

or as percent males:

$$\frac{\text{Total number of male offspring}}{\text{Total number of offspring born}} \times 100$$

These calculations should include both live and dead pups, since the sex of dead pups can usually be determined (unless they are partially cannibalized).

A treatment-related difference in the sex ratio values on lactation day 0 could be indicative of a deleterious reproductive effect, particularly if the skewed ratio is accompanied by an increase in preimplantation or postimplantation fetal loss. Alterations in sex ratio may be related to such factors as selective loss of male or female fetuses or selective impairment in the production, transport, or fertilizing ability of sperm which carry either the X- or Y-chromosome. Sex-linked genetic anomalies which are incompatible with life and which result from germinal mutations in the treated sire or dam, may also produce an altered sex ratio.

4. Offspring Clinical Observations

At the examination of litters following parturition, each pup is carefully checked for any visible external malformations or variations. Malformed pups left in the nesting cage will often disappear (due to maternal cannibalization) before the next observation interval. In addition, pups which are noted to be cold to the touch, weak, pale, or cyanotic are less likely to survive and may disappear from the litter count.

Bruises to the bodies, and even missing portions of appendages (such as the tail or foot), are common findings in newborn rodents. Such damage is generally the result of the birth process and/or rough treatment by the dam. The report should attempt to distinguish between missing or altered body parts which are due to malformation and those that are due to amputation.

Although the measurement of anogenital distance on lactation day 0 is not required by the 1982 FIFRA guideline, some studies will present this information. The protocol requirement for these data can be triggered for selected studies by the observation of treatment-related differences in anogenital distance on previously conducted developmental toxicity studies, or from structure-activity analysis or subchronic study findings (i.e., histopathological changes to the reproductive organs) that indicate the possibility that the test substance has an endocrine-mediated effect. Reduced fertility in the F1 generation and/or histopathological alterations in the reproductive organs of the F1 adults may provide confirmatory evidence for apparent treatment-related changes in anogenital distance. Vinclozolin is an example of an antiandrogenic substance which produced reduced anogenital distance (and other abnormalities of the reproductive organs) in F1 male rat pups (Gray et al., 1993).

As the pups grow and develop, observations of other abnormalities which were not readily seen in the newborn may be reported, i.e., evidence of microophthalmia may not be observed until the eyes of an affected pup are open, slightly domed skulls which are indicative of hydrocephaly may not be noticable until an affected pup is several days old, development of nipples in endocrine-altered male pups may not be observed until sometime between days 11 and 14 of lactation (e.g., vinclozolin, Gray et al., 1993) and problems with locomotion may not be apparent until the pups begin to leave the nest on their own and travel about the cage. In addition, indications of test substance toxicity may manifest themselves following ingestion of the treated diet by older pups (see Section IV. B.7.).

Since it is unusual for pups to be individually identified on any two-generation study, tracking specific pups through the clinical observation data may be impossible. The data are generally reported as summary incidences of observations at standard time intervals (lactation days 0, 4, 7, 14, and 21). Occasionally, incidences of observations which indicate reduced vitality, e.g., cold, weak, prostate, cyanotic, or pale, are appropriately grouped together for analysis. The reviewer should evaluate the summarized pup clinical observation incidences in the context of the concurrent offspring survival data and mean body weight values.

5. Pup Body Weights

In addition to the above offspring indices, a very important measurement of reproductive toxicity is the weight of the surviving pups. The 1982 FIFRA Guidelines suggest that live pups should be counted and litters weighed, by weighing each individual pup (optional) at birth, or soon thereafter, and on days 4, 7 (optional), 14, and 21 after birth. Pups are generally not uniquely identified, even when they are weighed individually; therefore, it is usually not possible to track a body weight (and clinical observations) history for individual pups from the data presented.

Pup body weight data should be evaluated concurrently with pup survivability data. As discussed earlier, an increase in offspring mortality after postnatal day 4 may result from either nutritional deficiency (hormonal imbalance in the dams), maternal neglect (behavioral change), or directly from the toxicity of the chemical tested (excreted through the milk or ingested in feed). Therefore, an increase in offspring mortality without impairment of pup body weights may, at least, rule out the possibility of nutritional deficiency, whereas concurrent increases in offspring mortality and decreases in pup body

weights may result from any or all of these factors. Regardless of the exact etiology of increased mortality and decreased pup body weights, these findings are generally considered to be toxic effects. As with other toxicological parameters, if an inconsistent but statistically significant decrease in pup body weight is found in the treated groups, the data should be compared not only with the concurrent control but also with control data from other generations and, if available, recent historical control data.

It is also important to keep in mind that litter size has an important influence on pup weight. Mean pup weight shows a slight but consistent decrease with increasing litter size when litters number six pups or greater (Khera et al., 1989). If litter size is greater in treated groups than in the control group, whether due to chance or as a compound-related effect, decreased pup weights may be expected. The HED Ad hoc Committee for Atrazine Reproductive Issue (USEPA, 1992) discusses the setting of the NOEL in a reproduction study with varying mean litter sizes).

The weight of the pups at weaning (lactation day 21) is another important parameter that should be considered in the evaluation process. A difference in neonatal birth weight between control and treated pups does not necessarily imply that a difference in weaning weight will ensue. The weaning weight may be similar to controls, suggesting a reversible effect, or remain altered, suggesting an irreversible effect. However, attainment of expected weight at weaning does not demonstrate that untested functional effects have neither occurred nor persisted.

In late lactation, the pups become very active and begin to play in and eat the food presented to the dam. For treated groups, this results in additional exposure of the pups to the test substance. They are receiving it in the milk from nursing, in the food consumed, and even dermally from playing in the food container. This additional exposure to the pups may result in evidence of increased toxicity late in lactation, including treatment-related decreases in body weight, mortality, and adverse clinical findings.

6. Crown-rump Lengths

Crown-rump length measurements are not required by the 1982 FIFRA Guidelines. However, if available, these data are useful for evaluating offspring growth in conjunction with the body weight data. Crown-rump length measurements are usually positively correlated with body weight.

7. Developmental Milestones

The evaluation of developmental parameters in the study offspring, although not required by the 1982 FIFRA guidelines, can provide useful information regarding sexual maturation as well as indications of hormonally-mediated effects on the offspring. Examination of pups immediately after birth for alterations in anogenital distance may help identify early subtle signs of reproductive toxicity. At day 14 of lactation, examination of the pups for normal development of nipple structure would further aid in identification of feminized males. The onset of sexual maturity, generally a body weight-dependant milestone, can be evaluated by examination of females for the opening of the vaginal orifice (Adams et al., 1985), and by determination of preputial separation in the males (Korenbrot et al., 1977). Other developmental milestones, such as measurements (time-to-event) of eye opening, pinna unfolding, incisor eruption, piliation, righting reflex, cliff avoidance, and negative geotaxis, may also be included in some reports (Brunner et al., 1979; Butcher et al., 1984; Hard et al., 1975; Robertson et al., 1980). Delays in any of these normal processes, whether or not linked to body weight, may indicate a toxic response to treatment.

8. Necropsy and Histopathology of Offspring

The 1982 FIFRA Guidelines (§83-4) states that "a complete gross examination should be done on all animals, including those which died during the experiment or were killed in moribund conditions". Although this wording implies that all offspring should be necropsied, the National Research Council (1977) suggested that only 10 male and 10 female pups, randomly selected from the F2b litters of each test group and the control group, should be sacrificed at weaning and subjected to a complete gross examination. An alternative approach includes the random selection and necropsy of one pup per sex per litter of the F1 weanlings not destined to become parents of the second generation. In this alternate study design, F2 pups might not be examined grossly unless changes were observed in the reproductive organs of the necropsied F1 weanlings. In the absence of guideline recommendations that clarify this issue, all approaches are considered acceptable.

A selected list of effects related to the reproductive organs which should be looked for in necropsied pups is presented below (E. Gray, personal communication, 1993).

Males:

1. small testes-undescended, ectopic or abdominal small scrotal testes
2. small or missing epididymides
3. small or missing accessory sex glands

4. persistent Mullerian ducts-vaginal pouch
5. hypospadias
6. cleft phallus
7. tumors of testes

Females:

1. small ovaries--ovary atrophy due to loss of oocytes
2. oviductal abnormalities or agenesis
3. small uterus
4. uterine abnormalities, including agenesis, tumors
5. cervical and vaginal abnormalities including agenesis
6. pituitary adenomas
7. abnormalities of the mammary glands, agenesis or tumors
8. retained Wolffian ducts and derivatives

Neither organ weight nor histopathology data of the pups are mentioned in the FIFRA Guidelines (USEPA, 1982). One suggested approach to postlife evaluation of weanlings includes weighing the ovaries, testes, brain, liver, kidneys, and known target organs from those F1 pups selected for gross necropsy. The ovaries, testes, target organs, and grossly abnormal tissue would be preserved for histopathological examination. Organs demonstrating treatment-related changes in the weanlings would be examined histopathologically for the control and high-dose groups. In this approach, the F2 weanlings would not be routinely examined because they would not be expected to provide a source of new or different information. Again, due to the lack of guidance contained in §83-4, any approach to the collection of offspring organ weight and/or histopathology data will generally be considered acceptable.

C. Other Considerations in Study Interpretation

As in any toxicity study, the design of a multigeneration reproduction study has certain limitations.

1. Associating Effects with Gender

One of the most important limitations of the reproductive data is the inability to clearly identify male and female adverse reproductive effects. Both males and females are routinely treated in reproductive studies, and changes in reproductive indices reflect the contributions of both sexes. Although suggestions of gender-specific effects may arise from organ weight changes and histological examinations conducted in this study (or from evidence in other studies such as subchronic feeding or developmental toxicity), conclusive evidence of gender-specific toxicity requires testing beyond that routinely required by the Office of Pesticide Programs. A probable male-mediated contribution to an effect may be more easily shown if histopathology and sperm

evaluations are requisites of a study design. In the presence of probable male-mediated effects, a request for a "crossover mating" study, in which treated males are mated to untreated females, is sometimes necessary. However, risk management decisions by the EPA have increasingly become gender neutral, i.e., designed to protect both sexes, and determination of gender sensitivity thus may not be essential.

2. The Lack of Specificity and Sensitivity

It should also be borne in mind that many reproductive indices lack sensitivity and specifity. For example, parturition is not always well monitored since rodents, in general, deliver nocturnally. The litter size is only an estimate, since a correct litter size should include cannabalized pups, which are not readily accounted for in a reproduction study. The effects on germ cells, gametogenesis, libido, implantation, and embryonic growth and survival are only indirectly measured in the reproduction study (Schwetz et al., 1980).

The identification of adverse effects in the reproduction study focuses on infertility, adverse pregnancy outcomes, and adverse effects on offspring survival and growth. However, there are other aspects of reproductive toxicity which may also be of concern. These include (1) subtle alterations in structural or functional competence of the testes or ovaries, (2) feed back mechanisms, (3) onset of puberty, (4) vaginal cytology, (5) premature reproductive senescence, (6) histological evaluations of accessory sex glands and organs, (7) histological evaluations of spermatogenesis, (8) endocrine evaluations, (9) biochemical markers, and (10) evaluations of sexual behavior. These parameters are generally not included in a standard §83-4 two-generation reproduction study, but may be available in the open literature or from an Agency laboratory which devotes its resources to specialized testing.

3. Test System Limitations

Test system limitations include, but are not limited to, issues of poorly defined indices, limitations in the rat as a surrogate for treatment-related potential human reproductive effects, undetected alterations in functional reserve capacity, effects mediated through inheritance of recessive genes, and the necessity of historical control data to define the biological significance of developmental toxicity findings.

a. Poorly Defined Indices

Poorly defined indices often tend to cloud or bias findings (USEPA, 1987). Summary data can be useful but should always be submitted in conjunction with individual untransformed data sets to allow evaluations of individual end points. In addition, indices may have more than one definition, and the reviewer must carefully examine these to assure that a meaningful and consistent description is correctly presented in study reviews.

b. The Relevance of the Animal Model in the Detection of Effects on Fertility

The rat is not an ideal surrogate in studies which are designed to assess the potential for human reproductive effects. For example, the female rat is a poor model for luteal function, lacking the formation of functional corpora lutea unless pregnant, with prolactin as the luteotrophic hormone. In addition, reproductive senescence in the female rat is mediated in the CNS (the hypothalamus) and not by oocyte depletion, as in humans. The male rat is also a poor model, since for example, the ejaculate in the rat generally contains a large excess of spermatozoa and a small (or even a moderate to severe) reduction in sperm count would not necessarily be detected in mating trials. In contrast to rats, humans lack this large excess capacity. An 80-90% reduction in sperm count which has little or no effect on fertility in the rat will generally result in functional sterility in the human male (Johnson, 1986). There is the opportunity for repeated copulations during each cohabitation in the multigeneration reproduction study and although the likelihood of fertilization may be reduced for each mating, repeated matings may minimize the sensitivity of the study to detect adverse effects on fertilization. It has been suggested that limiting the length of the mating period may increase the sensitivity of the multigeneration reproduction study (Zenick and Clegg, 1989).

c. Inability to Detect Effects Upon Functional Reserve Capacity in Neonates

Perinatal and postnatal functional maturation could be partially impaired but remain undetected during clinical observations or through examination of indices assessing viability. For example, exposure to a toxicant during fetal stages may result in diminished respiratory reserve capacity which would only be detected in the conduct of special testing. The animal may look and function normally, but, when challenged, may show an abnormal response. Its lungs may not fully inflate, the septal walls may be inadequately attenuated, and oxygen consumption may be depressed (Johnson, 1986); Newman et al., 1983 and 1984). Adverse effects on reserve capacity after <u>in utero</u> exposure may also be observed on the renal or immunological systems (**Kavlock ?, etc.**). Effects such as these

may not be discovered in routine multigeneration reproduction studies.

d. Recessive Gene Effects

Palmer (1981) cites the following (Table 8) as an example of an increased incidence of abnormal offspring in F1 and F2 generations which was not due to treatment but rather to the inheritance of a recessive gene as seen through the careful examination of derivation records through the three generations. Although such information will not generally be available in most multigeneration reproduction studies, this example illustrates how a factor which is not readily apparent, such as genotype, can influence study results.

Table 8. Effects of a Recessive Gene in a Rat Multigeneration Study[a]

Treatment	P Generation	F1 Generation	F2 Generation
Control	0.0	2.6	0.0
Low	0.0	0.0	8.2
Intermediate	13.0	2.7	6.7
High	0.0	6.8	7.1

a Incidence (%) of litters containing offspring with locomotor incoordination.

The gene was only expressed in the intermediate dose level in the P generation, since it was only at that dose level that one or more matings of heterozygous parents occurred. Such matings occurred more frequently in the two subsequent generations due to chance, and the low background expression of this gene clearly does not reflect a compound-related effect. It is important to be aware of this possibility since such findings may not be readily apparent upon first inspection of the data.

4. Use of Historical Control Data

Reproductive performance varies widely within the same species due to genetic and environmental factors. Although concurrent control data are normally the most appropriate for comparison to data from treated groups, on occasion, historical control data for the same strain/species provide the reviewer with valuable information regarding the background rate for various reproductive parameters as well as normal variations and trends. In general, historical control data should not be used as a substitute for concurrent control results

and need not be routinely requested. However, testing laboratories should be encouraged to routinely collect historical control data.

The reasons for requesting historical control data for reproductive toxicity studies are similar to those for developmental toxicity studies:

a. Historical control data can provide a guide for determining the biological significance of statistically significant differences observed in a reproductive study. Such data may indicate whether a concurrent control group incidence(s), i.e., pregnancy rate or dams with 100% resorptions, are unusually low for the test species, thereby artificially enhancing the statistical significance of findings in treated groups, or whether nominally increased incidences for a treatment group are within the normal range seen for that strain/species. While the range of reported historical values is important, it is presumed that individual findings outside two standard deviations of the mean are outliers and therefore may not be acceptable for comparison against treatment groups in the interpretation of statistical versus biological significance.

b. Historical control data may also indicate trends in the overall vigor, fertility, or litter size of a particular strain/species which relate to genetic drift and can help in the interpretation of unusual findings.

Some specific examples of situations when historical control data would be necessary to clarify the findings in a reproduction study are presented by category.

a. The data contains unusual concurrent control findings, leading to difficulty in interpretation of the validity of significant findings in treated groups

Example 1: Neither a clear dose response nor a NOEL are evident for observations such as decreases in fertility, increased incidence of abortions or premature births, smaller live litters, or number of dead pup/litter; however, the values for one or more treatment groups are elevated above the concurrent control such that statistical or apparent biological significance is achieved.

Example 2: Fertility appears to be unusually low in the concurrent control and/or treatment groups and is suspected of impacting the overall validity of the study (see IV.A.9, IV.D.4.b. for further discussion on fertility).

b. There is a clear variation in response between generations

Example 1: Differences in various reproductive parameters, i.e., litter sizes or mean pup weights, are noted between control generations (F1, F2).

Example 2: If developmental landmark timeframes are recorded (i.e., anogenital distance, vaginal opening, or preputial separation), they may vary between generations, and this variation may appear to be biologically or statistically significant.

c. Unusual findings are evident in one or more groups

Example 1: Developmental anomalies or variations are noted in a treatment group but not the concurrent control, and the findings are not dose-related. Background incidence data from either reproduction or developmental toxicity studies with postnatal phases may assist in determining the biological significance of statistically significant findings.

Example 2: An apparently high incidence of an observation such as pup mortality is observed at all dose levels, including control, although no significance is demonstrated in treated groups.

Examples of the types of historical control data which might be necessary to effectively evaluate a reproduction study are given in Table 9. These represent the most critical information for a complete Agency interpretation of reproductive toxicity studies but are not a comprehensive listing of reproductive indices or study parameters. In order to facilitate study evaluation, the reviewer may request that other historical control data, not included in the table, be provided by the registrant.

For both developmental and reproductive toxicity studies with postnatal phases, it is preferable that the initiation of treatment for the first generation of studies which is used in compiling specific historical control data sets be within a range of ± two years of the in-life portion of the study under review (USEPA, 1985b; USEPA, 1993b). However, studies for which the in-life portion is within ± five years of the study under review are generally considered to be acceptable for inclusion into the historical control data set for reproduction studies. This differs from the criteria for developmental toxicity historical control data because the types of reproduction data requested (see Table 6) require more time to generate and the number of available studies is smaller. In addition, some reproductive parameters are not as sensitive to subtle changes in examination procedures or diagnostic criteria which may affect the type of fetal abnormalities (visceral and skeletal) recorded in developmental toxicity studies. When using historical control data, the reviewer should be aware that there will be some difference between animals of the same strain from different suppliers and that the extent of inbreeding of commonly used laboratory rats has resulted in certain long-term trends. The Charles River Sprague-Dawley rat, for example, has been heavier, shorter-lived, and more fecund (with larger litters). A wider range in allowable time frame for historical control data allows for potential genetic drifts in the strain to be included. Due to the very real possibility of genetic and prodedural drift, historical control data derived from studies conducted closer to the date of the study under review should be given more scientific weight than studies conducted at either end of the five-year time range limit. The use and interpretation of historical control information generated prior to the advent of GLPs may be complicated by a greater degree of variability than is currently observed.

Studies which are not representative or typical, either in study design and conduct, or in the condition (health) of the animals, should not be included in the historical data set. Individual study results as well as summary statistics, i.e., mean, median, standard deviation or error, and/or range values, must be included in the historical control data submitted to the Agency. Individual studies should be identified by dates of initiation and completion of the in-life phase, vehicles utilized (if any), method of administration, changes in feed or animal suppliers, environmental conditions, or other significant factors which change over time. An accurate description should be provided for animals used in each study represented in the historical control data set; this information should include species/strain/supplier/facility, age at time of mating, and any noteworthy patterns of mortality or other indications of disease. Histopathological data, if included, should include the names of the performing laboratory and pathologist. Data sets should be separated and identified by generation.

Table 9. Examples of Multigeneration Study Data Which Might be Included in a Historical Control Data Set[a]

General Information
Number of males and females mated
Number of females pregnant
Number of confirmed pregnancies/number of confirmed matings
Number of males and females surviving to scheduled termination
Number of females with abortions
Number of females with premature deliveries
Number pregnant up to parturition
Adult Reproduction Indices[b]
Copulatory Index
Mating Index
Fecundity Index
Male Fertility Index
Female Fertility Index
Parturition Index
Gestation Index
Pup Reproductive Indices and Parameters[b]
Live Birth Index
Sex Ratio
Live Litter Size Index
Sex Index
4-Day Survival Index (Viability Index)
21-Day Survival Index (Weaning Index[d 0-21]/Lactation Index [4-21])
Preweaning Index
Mean pup weight at:
Day 0 (birth)
Day 4 (pre-standardization)
Day 4 (post-standardization)
Day 7
Day 14
Day 21 (weaning)
Incidences of developmental anomalies

a Historical control data for each generation of a given study should be included. Adapted in part from Feussner et al., 1992.

b For definitions, see discussion on reproductive indices and parameters (Section IV.A.9; IV.B). Consistency between studies in calculation of indices is required. Actual numbers used in calculation of the indices should be presented.

Historical data sets derived from multigeneration reproduction and/or developmental toxicity studies conducted in multiple laboratories and published in the scientific literature (Clemens et al., 1992) may be considered by the reviewer in the interpretation of study results. These data may be particularly useful when the performing laboratory is unable to provide a historical control data set from their own facility, although such a situation is rare. Generally, however, these data must be regarded with a certain amount of caution, since there may be unknown variability in study conditions and conduct between various laboratories, as well as inapparent differences in technical procedures, interpretation of effects, and calculation of indices.

V. REPRODUCTIVE TOXICITY RISK ASSESSMENT

The following discussion attempts to place in perspective certain frequently encountered toxicological and exposure issues which must be considered in evaluating the reproductive toxicity potential of pesticide chemicals (Section A). In Section B, current and future risk characterization approaches are discussed.

A. Hazard/exposure Issues

The following section is intended to provide perspective for selected issues which often arise in performing a reproductive toxicity risk assessment and to offer guidance in addressing these issues.

1. Data Extrapolation Between Routes of Exposure

Reproductive toxicity data are primarily generated using dietary exposures (see Section VI.D.). Because pesticides are often found as residues on raw agricultural foods and related feeds, studies by this route are generally appropriate in the assessment of tolerances. Oral intubation (gavage) allows precise measurement of dosage and is sometimes performed with volatile chemicals. Dietary risk assessments are conducted with the oral NOEL for reproductive toxicity without concerns for route-to-route extrapolation. Dietary exposure may result in a different pattern of developmental effects from that observed with oral intubation (Giovanni et al., 1986). Studies of pesticides which are used as gases may be performed via inhalation. For more information on some general requirements for the inhalation route the reader is directed to the HED Standard Evaluation Procedures for inhalation studies (USEPA, 1988c).

Despite the fact that most exposure of applicators, mixer/loaders, bystanders and harvesters is by the dermal route, testing by this route is rare. A dermal absorption study is often submitted to clarify the potential dermal penetration of a substance in order to allow better estimates of the margin-of-exposure when the reproduction study was conducted via the oral route.

2. Postnatal Data: Neurotoxicity and Other Special Studies

Many pesticides, including those which are neurotoxic, must be evaluated for their potential effect upon the structure and functioning of the nervous system in offspring exposed during pregnancy and lactation (USEPA, 1991a).

In developmental neurotoxicity studies, the test substance is administered in the female rat from gestation day 6 through day 10 of lactation. Dosing (usually oral) is not performed on the day of parturition in animals which have not completely delivered their offspring. The neurotoxicity evaluation includes observations to detect neurologic and behavioral abnormalities, determination of motor activity, response to auditory startle, assessment of learning, neuropathological evaluation, and brain weights. This type of study may either be separate from an adult neurotoxicity study or be part of a multigeneration reproduction study. Further discussion of the history of postnatal neurotoxicological studies is found in the Guidelines for Developmental Toxicity Risk Assessment (USEPA, 1991b).

The need for postnatal evaluation should be expanded to include other systems than the nervous system when the compound has a hormonal action such as those discussed under Section B.3.b. below. In that case, a special protocol may be developed to evaluate endocrine-active agents.

3. Mechanism(s) of Action

Knowledge of the site and mechanism of action of observed reproductive/ developmental toxicity in test animal studies can either diminish or increase our

Table 10. Mechanisms of Action of Reproductive Toxicants[a]

Mechanism	Compound
Structural similarity	Steroid hormones Cimetidine Diethylstilbestrol Azathioprine 6-Mercaptopurine Halogenated polycyclic hydrocarbons
Chemical reactivity	Alkylating agents Cadmium Boron Lead Mercury
Metabolic activation	Ethanol Chlorcyclizine Dibromochloropropane Polycyclic aromatic hydrocarbons Cyclophosphamide Ethylene dibromide
Disrupted homeostasis	Salicylazosuphapyridine Halogenated polycyclic hydrocarbons Anticonvulsants Ethanol

a Adapted from Table 5 of Mattison and Thomford (1987) as modified from Mattison (1983, 1984).

concern for the human population. If the mechanism of toxicity is or is likely to be operative in human physiology, e.g., similar receptor site mechanims and similar metabolic activation route, then much greater weight should be placed on the experimental findings. General mechanisms of reproductive toxicity include direct acting toxicants (similar structures to endogenous chemicals, chemically reactive agents) and indirect acting compounds (altering hormonal control of the reproductive system, requiring metabolic activation, disrupting homeostasis). See Table 10 for specific chemical examples.

a. Mutagenicity, Carcinogenicity, and Cytotoxicity

Studies show that oocytes and sperm possess DNA repair capabilities (Lee and Dixon, 1978; Pederson and Mangia, 1978; Lee, 1983). In general, agents which are mutagenic/genotoxic may have a range of effects upon germ cells quite similar to that observed in somatic cells including (1) cell death, (2) incorporation and repair of mutations, or (3) incorporation and expression of mutations (Mattison and Thomford, 1987). Compounds which induce the latter forms of toxicity to germ cells may result in dominant lethality. Paternally-induced developmental toxicity has also been associated with certain compounds, e.g., lead and cyclophosphimide. A recent conference on male-mediated developmental toxicity explored the evidence for male-mediated developmental and reproductive toxicity (International Conference on Male-Mediated Developmental Toxicity, 1992). These effects have been suggested to be induced through a variety of mechanisms including (1) genetically-heritable alterations, (2) epigenetic alterations such as disruption of DNA methylation patterns of cytosine residues which may alter the normal developmental program of male germ cell differentiation and subsequently the pregnancy outcome (Trasler, 1992), (3) microinjection of toxicant via the ejaculate (Silbergeld, 1992), and (4) direct effects upon the epididymus which alters maturation of the sperm (Robaire, 1992). As the basis for these male-mediated effects becomes more clearly differentiated, such data will require reconsideration of the overall approach to testing for reproductive toxicity.

The fact that a compound is a carcinogen does not necessarily mean that the same chemical will be a reproductive/developmental toxicant (or vice versa). However, diethylstilbesterol stimulates estrogen receptor-containing tissues and increases the risk of vaginal adenosis (75%), vaginal adenocarcinoma (0.01%), and anomalies in males (25%) (Ruddon, 1990). Other chemicals such as ethylene dibromide and dibromochloropropane are directly genotoxic and induce both reproductive and carcinogenic effects through this mechanism.

Direct cytotoxic agents are more likely to produce reproductive/ developmental toxicity than noncytotoxic agents if the cytotoxic threshold dose is achieved (Jelovsek et al., 1990). While it can be argued that such an effect is due to a general systemic effect rather than a specific reproductive mechanism, rapidly dividing germ cells in the conceptus may result in reproductive toxicity at dose levels lower than those at which general systemic toxicity is observed.

b. Endocrine Alterations

Differentiation in the fetus of the mullerian ducts (embryonic tubes from which the oviducts, uterus and vagina develop) and Wolffian ducts (embryonic tubes from which the ductus deferens, ductus epididymis, seminal vesicle, ejaculatory duct, ureter and pelvis of kidney develop) into the female and male reproduction organs and accessory tissues occurs during the latter period of organogenesis (Taber, 1970; Williams, 1974). These embryonic tissues are significantly affected by the endocrine environment (androgen) which directs the subsequent development of the fetus into male or female offspring. Specifically, testosterone stimulates the primitive Wolffian ducts to differentiate but fails to effect involution of the Mullerian structures.

Compounds which have androgenic or anti-androgenic activity may affect normal sexual differentiation. Procymidone [N-(3,5-dichlorophenyl)-1,2-dimethylcyclopropane-1,2-dicarboximide] is a fungicide which has recently been shown to be negative when tested for developmental toxicity in two species (rat and rabbit) at high dose levels (up to 300 to 1000 mg/kg/day) (USEPA, 1990). However, in a two-generation reproduction study, dietary levels of 750 ppm produced reproductive and developmental toxicity including abnormalities of external genitalia (reduced anogenital distance and hypospadias). Although procymidone has been shown to have a low affinity for androgen receptors in prostate cytosol (0.07% of dihydro-testosterone), it is likely that the effects noted were mediated by a disturbance in endocrine function. Vinclozolin [3-(3,5-dichlorophenyl)-5-methyl-5-vinyl-1,3-oxazolidin-2,4-dion], a structural analog of procymidone, produces a similar pattern of developmental effects in male rats when administered either dermally or orally (gavage) in standard developmental toxicity studies during gestation days 6-19 but not when administered on days 6-15 of gestation (Hellwig, 1989).

4. Pharmacokinetic/Physiologic Considerations

For the same chemical, there are often species differences in both pharmacokinetics and activity at the target site. Comparative pharmacokinetic

data have rarely been available for pesticides but may explain much of the difference in the response of various species, strains and genders to test chemicals. Furthermore, with the advent of physiologically-based pharmacokinetic modeling, comparative pharmacokinetics may allow more accurate extrapolation between species (Gabrielson and Larson, date). In the absence of good predictive data regarding concentrations of the proximate toxicant at the target site, testes, ovary and embryo, in man and in the test species, it is generally assumed that humans and test species are similar in the pharmacokinetic disposition of the chemical and in the response of the target tissue. Differences in species sensitivity at the target site have been documented less frequently but may not be an important factor for chemicals with steroidogenic properties. However, most differences are pharmacokinetic in nature and not pharmacodynamic.

Pharmacokinetic studies/modeling are assuming a much greater risk assessment role within the EPA by linking exposure with developmental effects observed in the fetus (animal, humans), and being utilized to properly design developmental toxicity studies. Tables 11 and 12 are presented below with information useful to pharmacokinetic considerations. The reader is directed to the recent symposium and reviews on pharmacokinetics in developmental and reproductive toxicity for further reading on approaches being taken (Kavlock, 1991; Nau and Scott, 1987; Ribeiro and Faustman, 1989).

a. Comparative Pharmacokinetics

Significant pharmacokinetic differences between humans and animals are evident (Nau, 1991). It is important to note that the half life of xenobiotics are often an order of magnitude shorter in experimental animals than in humans. During conventional developmental toxicity studies, steep concentration-time peaks are often produced due to rapid absorption and elimination; these high peaks rapidly fall to low levels. However, target sites in the human may be exposed to the toxicant for longer periods of time due to the longer half-lives. Furthermore, the first-pass effect (rapid metabolism in the liver due to the direct transport of orally administered chemicals from the gut via the portal vein) is sometimes more extensive in animals than in humans. Human maternal plasma protein binding is often more extensive than in the plasma of experimental animals (Nau, 1991). Finally, the duration of the sensitivity of particular developmental processes are often several-fold shorter in experimental animals than in man. Thus, in experimental studies a multiple-dosing regimen during a defined period may be more useful than the conventional once-daily administration regimen, where the time of maximal sensitivity may be missed.

Table 11. Body Weights, Surface Areas, and Conversion Factors of Dosing from mg/kg into mg/m^2 [a,b]

Species	Body Wt. (kg)	Surface Area (m^2)	Conversion Factor	Dose Equivalent (kg)[c]
Mouse	0.02	0.0066	3.0	12.0
Rat	0.15	0.025	5.9	6.0
Dog	8	0.40	20	1.7
Monkey	3	0.24	12	3.0
Human				
Child	20	0.80	25	1.5
Adult	60	1.60	37	1.0

a Nau and Scott (1987), p.95.
b To convert a mg/kg dose in a given species into an equivalent mg/m^2 dose, the dose is multiplied by the conversion factor.
c Dose equivalent for the adult human is set as 1.0.

Table 12. Physiological Characteristics of Various Species Relevant for Pharmacokinetics[a]

Physiological Characteristics	Species						
	Mouse	Rat	Guinea Pig	Rabbit	Dog	Monkey	Man
Bile flow (ml/kg x day)	100	90	230	120	12	25	5
Urine flow (ml/kg x day)	50	200	---	60	30	75	20
Cardiac output (ml/min x kg)	300	200	---	150	100	80-300	60-100
Hepatic blood flow (L/min)	0.003	0.017	0.021	0.12	0.68	0.25	1.8
(ml/min x kg)	120	100	---	50	25	25	25-30
Liver weight (% of body wt.)	5.1	4.0	4.6	4.8	2.9	3.3	2.4
Renal blood flow (ml/min x kg)	30	---	---	---	22	25	17
Kidney clearance (ml/min x kg)	5	---	---	---	3.2	3	1.3

a Nau and Scott (1987), p. 95.

Reproductive toxicity (single or multigeneration) tests are preponderantly dietary in nature, and, therefore, the peak plasma concentrations are usually lower and of longer duration than after an oral bolus of the same compound. This is due to frequent food consumption in rodents; the slower absorption of compound from the gastrointestinal tract due to the presence of the food would yield a more persistent presence in the blood. However, more extensive metabolism and/or binding to plasma proteins may occur with lower plasma concentrations due to the saturation of binding capacity of metabolic pathways at higher plasma concentrations. For significant inhalation or dermal exposures associated with workers, where the likelihood of much higher peak plasma concentrations is present, the standard reproduction test may not be the optional study design to address the reproductive potential of a pesticide, and special studies may be required.

b. Physiological Alterations During Pregnancy

During pregnancy physiological changes in several systems can alter the pharmacokinetics in both the mother and fetus (see Table 13, from Mattison et al., 1991). These physiological alterations are required for successful pregnancy and lactation and result from maternal homeostatic mechanisms to deliver essential nutrients to the fetus and remove heat, carbon dioxide, and waste products from the fetus. These alterations are species dependent (e.g., cardiac output is increased 50% in humans during pregnancy, 20% in rabbits) and may involve different physiological strategies.

Two major alterations in xenobiotic elimination are renal and extrarenal elimination pathways. These pathways should be kept in mind while reviewing the pharmacokinetic/metabolic nature of the agent under consideration.

i. Renal Mechanisms

Renal function probably undergoes the greatest physiological changes during pregnancy (Krauer, 1987). Based on the fact that renal plasma flow and glomerular filtration almost double in humans during pregnancy (Davison and Hytten, 1974, as cited in Krauer, 1987), it may be inferred that for xenobiotics which are eliminated predominantly in the urine and are not highly protein bound, plasma concentration is generally lowered, half-life decreased and clearance increased in parallel to the increased renal function (see Table 14 for examples of changes in kinetic parameters).

Table 13. General Physiological Changes During Pregnancy[a]

Parameter	Change
Absorption	
Gastric emptying time	Increased
Intestinal motility	Decreased
Pulmonary function	Increased
Cardiac output	Increased
Blood flow to skin	Increased
Distribution	
Plasma volume	Increased
Total body water	Increased
Plasma protein	Decreased
Body fat	Increased
Metabolism	
Hepatic metabolism	±
Extrahepatic metabolism	±
Plasma proteins	Decreased
Excretion	
Renal blood flow	Increased
Glomerular filtration rate	Increased
Pulmonary function	Increased
Plasma proteins	Decreased

a From Mattison et al., 1991.

Table 14. Kinetic Parameters of Xenobiotics with Predominantly Renal Elimination ($Q_o \leq 0.3$)[a]

Drug	Q_o	Protein binding (%)	Changes in kinetic parameters			
			C_p	V_d	$t_{1/2}$	Cl_{tot}
Ampicillin	0.1	15-29	↓	↑	↓	↑
Cephacetrile	0.04	23-26	↓		↓	
Cephalexin	0.04	15	↓			
Cephazolin	0.06	84	↓		↓	
Cefuroxime	0.07	40	↓	↑	↓	↑
Digoxin	0.3	20-40	↓			↑
Kanamycin	0.03	0.3	↓			
Lithium	0.02	0	↓			↑
Sotalol	0.1	54	↓			↑

a From Krauer, 1987.
Q_o = the extrarenal dose fraction which indicates the percentage of the absorbed dose not excreted unchanged in the urine.
C_p = plasma concentration; V_d = volume of distribution; $t_{1/2}$ = half life; Cl_{tot} = total clearance.

ii. Extrarenal Elimination (Liver)

The most important extrarenal elimination occurs in the liver and is dependent on hepatic blood flow, the capability to metabolize drugs (intrinsic hepatic clearance) and the extent of binding in plasma (Krauer, 1987). Xenobiotic disposition varies and can be assessed from the magnitude of the hepatic extraction ratio. In human pregnancy and labor, intrinsic hepatic metabolism may be altered. Changes in xenobiotic protein binding have been shown to be quite significant during pregnancy (Krauer, 1987). Changes in plasma xenobiotic free fraction will only be relevant for those chemicals extensively (>85 to 90%) bound. However, from studies with many different drugs (chlorazepate, etidocaine, labetalol, meperidine, metronidazole, oxazepam, phenobarbitone, phenytoin, propanolol, thiopental, valproate, caffeine, diazepam, metoprolol) it is apparent that no general rule can be derived regarding agents primarily eliminated via the hepatic route.

7. Structure-Activity Relationships

Structure-activity relationships have only been studied to a limited extent for reproductive/developmental toxicants. Tables 15a and 15b present a list of known or potential human reproductive toxicants which have been associated with certain sites of action within the reproductive process. This list, while not comprehensive, may be useful to identify potential structural analogs for chemical agents under consideration for Peer Review. A more complete listing of inferred reproductive toxicants based upon animal data is given in Hayes (1982).

Certain agents with known actions such as hormonal activity, alkylating ability, CNS/peripheral nervous system activity, or microtubule disruption, and which are highly lipophilic in nature should be suspect due to the obvious susceptibility of the reproductive process to perturbation by the actions of such agents.

8. Human Data

In the area of pesticide toxicity, human data are primarily available for establishing exposure rather than establishing reproductive hazard. Such data may be derived from acute poisoning cases or biological monitoring (urine, blood, dermal patches) of field workers or mixer-applicators, in state, county or registrant-sponsored studies. Exposure estimates may also be obtained by surrogate analyses.

Table 15a. Known or Potential Male Reproductive Toxicants[a]

Site of Action	Examples
Endocrine System	
Anabolic steroids....	
Antiandrogens.....	(cyproterone acetate, spironolactone, cimetidine)
Estrogens/estrogenic.	(estradiol-17ß, DES, Estracyt, DDT, methoxychlor, chlordecone, mirex, PCBs)
Organochlorine.....	(dieldrin, aldrin, chlordane)
Miscellaneous.....	(cadmium, ethylene dibromide, clofibrate, alcohol, marihuana, morphine, methadone, phenytoin, primidone, phenobarbital, dibromochloropropane, borax, carbaryl, carbon disulfide)
Blood-testes barrier, sertoli cell function, spermatogenesis	
Antineoplastic agents	(chlorambucil, cyclophosphamide, busulphan, vinblastine, cytosine arabinoside)
Pesticides.....	(benomyl, carbaryl, Ordram, ethylene dibromide, dibromochlorpropane, diquat, paraquat, maneb, zineb)
Glycol ethers.....	(ethylene glycol monomethyl ether, ethylene glycol monomethyl ether acetate, ethylene glycol monoethyl ether, ethylene monoethyl ether acetate, diethylene glycol dimethyl ether, diethylene glycol monoethyl ether)
Phthalate esters.....	(di-2-ethylhexylphthalate, mono-2-ethylhexylphthalate)
Metals.....	(cadmium, borax, manganese chloride, lead)
Phenoxy herbicides...	(2,4,5-trichlorphenoxyacetic acid, 2-methyl-4-chlorphenoxyacetic acid)
Industrial agents....	(carbon disulfide, 2,3,7,8-tetrachlorodibenzo-p-dioxin/TCDD, chloroprene, Tris, carbon tetrachloride)
Miscellaneous.....	(acetominophen)
Sperm maturation, epididymal function	
Fungicides, solvents.	(dibromochloropropane, epichlorhydrin, ethylene dibromide)
Metals.....	(cadmium)
Accessory glands, secretions	
Miscellaneous.....	(methadone, CCl_4, methoxychlor, EDB, DDT, PCBs, chlordane, dieldrin, lead, cadmium, manganese salts, 2,4,5-T, MEHP)
Ejaculatory process	
Antihypertensive.....	(hexamethonium, pentolinium, chlorisondamine, mecamylamine, pempidine, bretylium, guanethidine, reserpine, clomidine, methy-dopa, propanolol)
Antipsychotics.....	(phenothiazine-- thoridazine, chloropromazine, thiothixene, haloperidol, fluphenazine, trifluoperazine, butaperazine, chlorprothixene, piperacetazine, mesoridazine, molinodone, perphenazine, triflupromazine)
Other CNS agents.....	(tricyclic antidepressants - imipramine, protryptyline, desmethylimipramine, amitriptyline, clomipramine; anticholinergic - methamtheline; monamine oxidase inhibitors; lithium; nitrous oxide; hexachlorophene)
CNS depressants.....	(barbiturates, methaqualone, alcohol)
CNS stimulants.....	(amphetamine)
Fertilization process	
Miscellaneous.....	(alcohol)

a From Waller et al. (1985); many of these are confirmed in human studies.

Table 15b. Known or Potential Female Reproductive Toxicants[a]

Site of Action	Examples
Endocrine System	
Estrogens/estrogenic	(mestranol, norethindrone, DES, DDT, PBBs)
Industrial chemical	(2,3,7,8-TCDD)
Menstrual (Estrous)/CNS Effects	
Metals	(lead, mercury)
Pesticides	(carbaryl, DDT, EDB[?])
Industrial chemicals	(2,3,7,8-TCDD, PBBs, PCBs, formaldehyde, chlorprene)
Organic solvents	(carbon disulfide, benzene, styrene)
Ovarian: Direct Damage, Accumulation, Impaired Function	
Metals	(mercury, boron)
Pesticides	(carbaryl, DBCP, DDT, chlordecone)
Industrial chemicals	(2,3,7,8-TCDD, PCBs, 1,3-butadiene)
Organic solvents	(benzene, carbon tetrachloride)
Placental Effects	
Metals	(mercury, cadmium)
Industrial chemicals	(PCBs)
Spontaneous Abortions/Stillbirths	
Metals	(lead, mercury, arsenic)
Industrial chemicals	(PCBs, ethylene oxide, rubber chemicals)
Organic solvents	(carbon disulfide, styrene (?))
Anesthetics	(nitrous oxide)
Delayed Partuition	
Industrial chemicals	(PCBs)
Postnatal: Low Birth Weights/Slow Weight Gain/Behavioral/Learning/Death	
Metals	(lead, manganese, mercury, cadmium, antimony)
Industrial chemicals	(PBBs, PCBs)
Impaired Fertility of Offspring/Tumors	
Estrogens	(DES)
Agricultural chemicals	(DDT)
Industrial chemicals	(PBBs, PCBs, solvents)
Breast Milk Concentration	
Agricultural chemicals	(DDT, chlordecone)
Industrial chemicals	(PBBs, PCBs)
Other: "Decreased Fertility"/Sterility (Mother)	
Estrogens	(DES)
Metals	(lead, boron)
Agricultural chemicals	(DBCP, DDT)

a From U.S. Congress OTA (1988); many of these are confirmed in human studies.

9. Measurements of Additional Endpoints Not Currently Required by FIFRA Guidelines

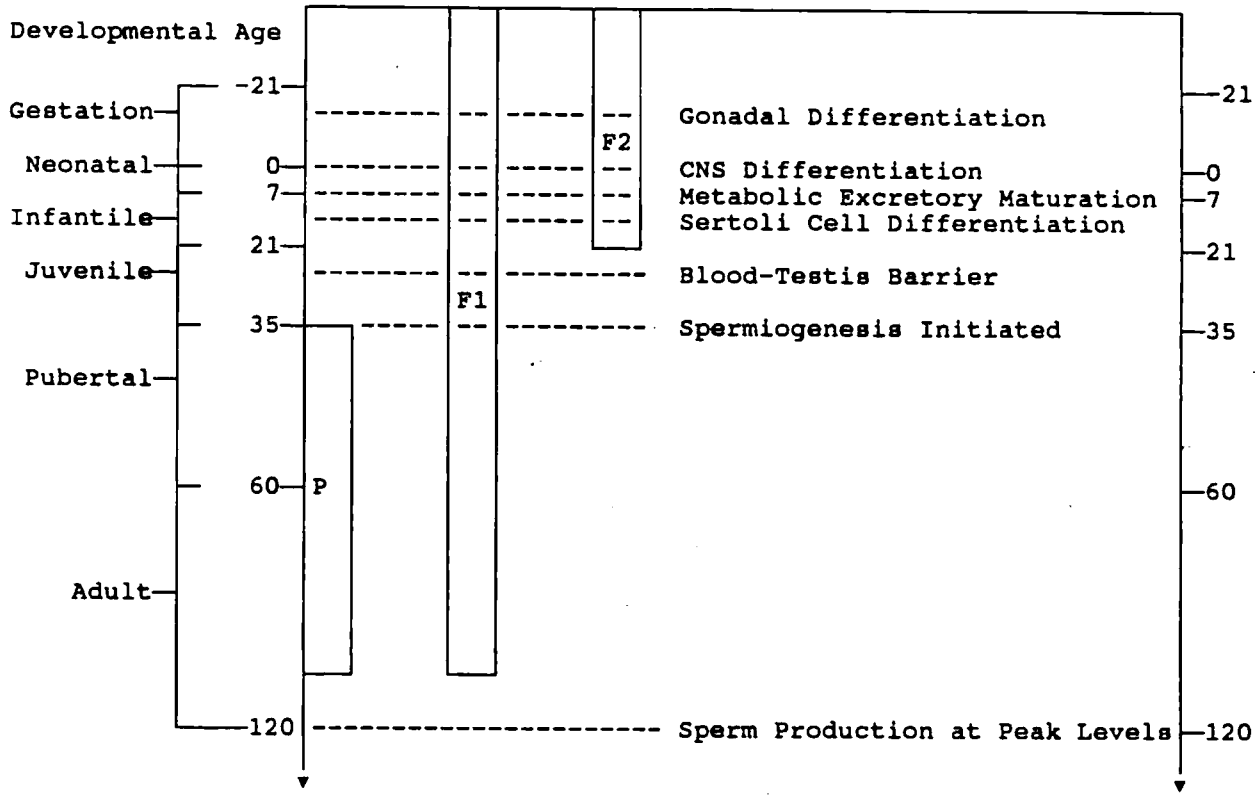
There is an ongoing reexamination of the adequacy of the reproductive toxicity testing protocols both at the Program and Agency level to evaluate potential male and female reproductive toxicity (USEPA, 1993a). A summary of potential endpoints/markers for incorporation possibly in (1) experimental animal subchronic, reproductive, or special studies or (2) direct human assessments is presented below. It is likely that at least some of these endpoints will find their way into future evaluations of reproductive testing and will allow a more comprehensive assessment of the ability of a chemical agent to alter reproductive physiology.

a. Male endpoints

Recent laboratory and epidemiologic data suggest that chemical-induced effects upon the offspring (e.g., birth defects, cancer, or death of the conceptus), may be mediated in some exposures, through an effect upon the male alone (International Conference on Male-Mediated Developmental Toxicity, 1992). These effects have been suggested to be mediated by a variety of mechanisms including (1) genetically-heritable alterations (mutations) (Russell et al., 1990), (2) epigenetic alterations such as disruption of DNA methylation patterns of cytosine residues which may alter the normal developmental program of male germ cell differentiation and subsequently the pregnancy outcome (Trasler, 1992), (3) microinjection of toxicant via the ejaculate (presentation by Silbergeld, 1992), and (4) direct effects upon the epididymis which alters maturation of sperm (presentation by Robaire, 1992). As the basis for these male-mediated effects becomes more clearly differentiated, such data will require reconsideration of the overall approach to testing reproductive toxicity. For example, if a particular cell type in spermatogenesis is known to be sensitive to the test compound, it may not be necessary to expose males to the chemical for a full ten week period prior to cohabitation. Figure 2 below presents an over view of the developmental stages and events (primarily from a male perspective) during which treatment-related exposures occur in a two-generation reproduction study.

The biological markers presented in Table 16 are for assessment of physiological or genetic damage in human males but generally have their counterpart in animal testing.

Figure 2. Developmental Stages and Events Encompassed by Exposure of P, F1, and F2 Animals in a Two-Generation Reproduction Study[a]



a Zenick and Clegg, 1989.

b. Female endpoints

The biological markers presented in Table 17 are for human female assessment of physiological or genetic damage but generally have a counterpart in animal testing.

B. Risk Characterization

Risk characterization is the culmination of hazard assessment/dose-response and exposure assessment.

Table 16. Male Endpoints

Tissue or data	Markers of
Physiologic Endpoints[a]	
Testis (or biopsy)	Histopathology
Seminal sperm	Sperm number Structure Motility Double F bodies Viability Agglutination Penetration and egg interaction Cervical mucus Hamster eggs Nonliving human eggs Internal and surface domains Chromatin structure
Other seminal parameters	Physical characteristics Immature germ cells Non-germ cells Chemical composition Normal and xenobiotic constituents Sertoli cell, Leydig cell, and accessory gland function
Blood	Hormone levels
Survey and medical records	Fertility status Standardized fertility ratio Time to conception
Maternal urine	Indicators of early pregnancy
Genetic Endpoints[b]	
Testis	Cytogenetic analyses of cells in mitosis, meiosis I, and meiosis II
Semen - Sperm	Sperm cytogenetics Sperm DNA and protein adduction Gene mutations in sperm Sperm aneuploidy
Semen - Immature germ cells	Spermatid micronuclei Cytogenetics of ejaculated meiotic I cells
Questionnaire and medical records	Sex ratio Spontaneous abortion Offspring cancer Sentinel phenotypes
Offspring tissue	Cytogenetics DNA sequencing Protein mutations Restriction-length polymorphism of DNS RNAase digestion Subtractive hybridization of DNA Denaturing gel electrophoresis of DNA Pulse-field electrophoresis of DNA
Maternal urine	Detection of early fetal loss
Somatic cell surrogates	HGPRT mutations (in WBCs) Hemoglobin mutations (in RBCs) Glycophorin A mutations (in RBCs)

a From Table 7-1, Biologic markers of physiologic damage to human male reproduction, reviewed in Chapter 7 (NRC, 1989).
b From Table 9-1, Potential markers of genetic damage and heritable mutations in the male germline, reviewed in Chapter 9 (NRC, 1989).

Table 17. Female Endpoints[a]

Site	Marker
Exposure: Chemical analyses for toxicants or metabolites, or mutagenic analysis of body fluids	Blood, urine, saliva Tissues Intact Cytologic specimens Fluids Cerebrospinal fluid Follicular fluid, amniotic fluid Placental tissue Peritoneal fluid
Genotoxic - DNA adducts (chemical specific, generic)	Oocytes, ovarian tissue Placental tissue Fetal tissues Maternal serum Fetal serum Unscheduled DNA synthesis Maternal lymphocytes Fetal lymphocytes SCE (sister-chromatid exchange) Maternal lymphocytes Fetal cells Chromosomal abberations Maternal serum Abortus tissue Chorionic villi Amniotic cells Fetal serum Micronuclei Maternal blood Vaginal/cervical cells Fetal liver cells Fetal lymphocytes Specific-locus mutations
Development/aging	Onset of puberty Breast bud development Blood Melatonin DHEA-S Gonadotropin (pulsatile) Age of first menstrual bleeding Hormones: estrogens, inhibin, LH, FSH, androgens Age of breast development Sexual behavior Neurotransmitter in CSF Menstrual cycle length Ovarian-oocyte stock Ultrasound for ovarian size IVF Biopsy MRI Periodic ultrasound to monitor follicular development Inhibin Premenopausal hormonal status (estrogens, gonadotropins, inhibin, LH, FSH) CNS reproductive senescence

Table 17. Female Endpoints[a] - continued

Site	Marker
Menstrual function	Cycle frequency and characteristics Detection of corpus luteum Follicular development (ultrasound) Basal body temperature Thermometer Improved, self-recording electronic thermometer Cervical mucus Sexual behavior Vaginal cytology Biophysical measurements of vaginal secretions Endometrial histology Endocrinology: gonadotropins, steroids, ovulatory hormones In vitro assays Pituitary cells (from cadavers) Granulosa cells Luteal-specific proteins, endometrial cell cultures Mucus production, endocervical cells
Fertilization, implantation, and loss	hCG (human chorionic gonadotropin) EPF (early pregnancy factor) PEP (progestin-associated endometrial protein)

a From Table 16-1, Status of current and potential markers in female reproductive toxicology, Chapter 16 (NRC, 1989).

1. Reference Dose vs. Margins-of-Exposure (Margins-of-Safety)

a. Use of the Reproductive No-Observed-Effect-Level (NOEL)

Current practices within the Agency are to characterize developmental but not reproductive risk by using margins-of-exposure (MOE) (formerly known as margins-of-safety) or uncertainty factors (see Section V.B.1.b. below). The MOE is a direct comparison (ratio) between the appropriate No-Observed-Effect Level (NOEL) and the estimated human exposure. The uncertainty factor approach generally includes a 10-fold factor for interspecies variation and a 10-fold factor for intraspecies variation. Approaches for the generation of benchmark dose levels, are under development from models which utilize data at all dose levels and may be applicable to reproduction and developmental studies in the future.

NOELs from reproduction studies are considered along with systemic toxicity NOELs derived from other data sets (e.g., subchronic and chronic studies in dogs and rodents) in selecting the appropriate study for setting the reference dose (RfD) for a pesticide. Determining what constitutes a selective reproductive effect as opposed to a general systemic effect is a matter of careful analysis

of the parental and offspring effects and may not be easily resolved nor might it necessarily be of any importance in making a final toxicological decision. It is important to compare the resultant NOELs observed in reproduction studies against other long-term studies to determine if the pregnant animal is more sensitive than the non-pregnant female.

The NOEL from the most sensitive species tested (where multiple tests are available) is generally used for reproductive toxicity risk assessment purposes due to the great difficulty in determining the most relevant species from which to extrapolate to humans.

b. Subchronic/Chronic vs. Short-Term Exposures

In most instances, exposure assessment in OPP is framed differently for reproductive risk as opposed to developmental risk. This is due to the potential that even single acute/short-term exposures may result in a developmental insult whereas reproductive toxicity studies are generally associated with subchronic to chronic dietary exposure. It is recognized that this approach is an artifact of the testing procedures. Thus, for reproductive risk assessment, the use of NOELs from a reproduction test would not be generally compared against acute/repeated exposure situations such as those observed with mixer-loader-applicators. Rather, the comparison would be between the subchronic/chronic average daily exposure in the test species as compared to the potential or observed subchronic/chronic exposure in the human population of concern.

There are many reasons, related to the physiology of the reproductive system, the presence of highly sensitive individuals within an exposed population, and the nature of the individual chemical agent, which support the possibility that a short-term exposure (acute, high exposure; repeated moderate exposure; acute exposure to a potent reproductive toxicant) may be sufficient to produce reproductive toxicity of either a reversible or irreversible nature. For example, destruction of the Sertoli cells or spermatogonia in the testes or oocytes in the ovaries is an irreversible phenomenon which may occur from a single exposure and may permanently affect the reproductive capacity of the exposed individual. After ovulation, single exposures to compounds such as carbendizim also alter the fertilizability of the ova (Darney, 1990). Such an effect could cause significant delays in the ability of an individual to conceive, particularly one with low fertility potential. Determination that a pesticide has such acute reproductive toxicity potential must be done on a case-by-case basis.

c. Forms of Exposure: Dietary, Occupational, Drinking Water

Reproductive toxicity risk assessment should include dietary and worker exposure, as well as other forms of exposure such as drinking water or home use. Worker exposure estimates are the responsibility of the Occupational and Residential Exposure Branch (OREB). OREB estimates of exposure are on a daily basis and quantified for each route of exposure.

It is the responsibility of the OPP scientists to determine the rate of dermal absorption. If available, pharmacokinetic data such as peak plasma concentrations, area under the curve of the test material, and/or metabolites should be compared when dosing is by different routes. Metabolism data should also be considered, since this may vary with route of exposure. In the absence of dermal absorption data, a 100% rate of absorption is generally assumed. The reviewer may need to request a dermal absorption study to define the rate of absorption.

Dietary risks should be assessed using the Dietary Residue Exposure System (DRES). This system compares NOELs to the predicted dietary exposure for the appropriate subgroups (e.g., adult males, pregnant females, infants, and children) and for the period of exposure (acute or chronic) which is most appropriate for the form of toxicity which is the basis for the NOEL.

Drinking water risks may also be of concern and are assessed in a manner similar to dietary risk. Determination of whether or not a pesticide has a potential for groundwater or surface water contamination is the responsibility of the Environmental Fate and Effects Division (EFED). If actual contamination exists, the most relevant contamination levels must be selected in consultation with the EFED. The National Academy of Sciences has recommended that, for risk assessment purposes, it be assumed that the average adult consumes two liters of water per day (NRC, 1977; US EPA, 1986c; USEPA, 1989a). The estimated daily exposure (mg/kg/day) to a pesticide in drinking water is therefore determined by multiplying the appropriate estimate of the residue level (mg/liter) by two liters and dividing that amount by body weight. Further descriptions of the hazard evaluation of pesticides in drinking water are available in the literature (USEPA, 1989a).

2. Limitations of Non-multigeneration Studies

In reviewing reproductive data, the reviewer may encounter ancillary studies which do not fulfill the regulatory requirements for a complete two-

generation reproduction study. Nevertheless, these studies often provide important qualitative data supporting the potential reproductive toxicity of a chemical or provide support for the requirement for a full study or special studies to define possible mechanisms of toxicity. These studies and the major limitations are listed below in Table 18.

Table 18. Limitations of Other Reproductive Data (Ancillary)

TYPE OF REPRODUCTIVE STUDY	LIMITATIONS IN REGARD TO THE MULTIGENERATION STUDY
FDA Segment I	One generation study; variable premating exposure of males alone (60-80 days), males (60-80 days) and females (2 weeks)
Segment II	Developmental toxicity study; exposure of females during organogenesis (gd6-15, rats)
Segment III	Exposure limited to final one-third of gestation through weaning; one-generation study
Continuous Breeding Protocol	Original protocol involves continuous breeding of P males and females (one generation; multiple litters up to 5) for 14 weeks; limited premating exposure of both sexes (7 days)
SIDS	One generation; limited number of animals (10/sex); limited exposure (beginning 2 weeks premating); limited reproductive and histological data requirements
Dominant Lethal	Only males are exposed (one day to one week); limited examination of in utero implantations/resorptions at mid gestation; no lactation phase

VI. WRITING STUDY REVIEWS AND SUPPORT OTHER DOCUMENTS

Once data have been evaluated for acceptability and interpreted using the principles described in previous sections of this Standard Evaluation Procedure, the results must be described and summarized in support documents. The most basic of these documents is the Data Evaluation Record, commonly referred to as the DER, which provides information used in risk assessments. Section A, below, discusses the information that should be included in a DER, without regard to a specific format or style. Section B outlines the information necessary for presentation to the Health Effects Division Reproductive Toxicity Peer Review Committee (PRC). This approach is similar to that utilized for developmental

toxicity (USEPA, 1993b) and is consistent with the Agency guidelines for the assessment of reproductive toxicity risk (USEPA, 1993a).

A. The Data Evaluation Record

Each study DER is written by an Agency representative, under the direction of one of the Toxicology Branches of HED. The DER should adhere to the following format, as applicable.

The DER should be considered a scientific document that is consistent with accepted methods of technical writing. For example, the Instructions to Contributors which are included in all scientific journals can provide adequate instructions regarding format and presentation. The reviewer should refer to the 1982 FIFRA test guidelines as well as the Guidelines for Reproductive Toxicity Risk Assessment (USEPA, 1993a) when preparing the DER.

1. Cover Sheet

The first cover page of the DER should contain the following information:

a. **Identification of HED staff** who conducted the primary and secondary reviews by name, section and branch;

b. **Identification of study type** by name (two-generation reproductive toxicity study), species tested, and guideline number;

c. **Identification of the test chemical** by name (including synonyms), composition, structure (when possible), EPA Pesticide Chemical Code (Shaughnessy number), HED Chemical Number (Caswell number, Tox. Chem. No., etc.), and/or EPA Registration Number;

d. **EPA identification of the study report** by Master Record Identification Number (MRID No.) and Data Submission Number (numbers given a "D" or "S" prefix on the documentation accompanying the data package);

e. **Study references**, including author(s), title, testing laboratory, sponsor and/or submitter, study or report numbers assigned by the sponsor and conducting laboratory, and date issued;

f. **Summary and conclusion**, containing the dose levels tested, the strain of animal used, duration of dosing, route of administration, no-observed-effect level(s) (NOELs), lowest-observed-effect level(s) (LOELs) with a brief description of the effects and how they changed with dose; NOEL and LOEL values should be expressed in mg/kg/day separately by sex and dose level (see Sections III. 5. and IV. 4.);

g. **Evaluation of study acceptability**, which should (1) state whether the study satisfies the §83-4 requirement for registration, (2) describe deficiencies in the study and state whether they can be rectified by submission of additional information or conducting another study, and (3) classify the study (assign a "Core classification" of Guideline, Minimum, Supplementary, or Invalid) (see Section V.B.).

h. **Compliance**, which should indicate whether the report included signed and dated statements of confidentiality, compliance with GLPs, quality assurance review, or flagging criteria (1992 FIFRA §6(a)(2)

criteria, as described in 40 CFR 158.34).

Much of the information from the DER coversheet is used to prepare a "Toxicology One-Liner" for a database maintained by the Toxicology Branches. A "One-Liner" for a multigeneration study includes the following information from the above list:

Citation	Material	MRID No.	Results	Core grade Document No.
Guideline number, Study type, Species, Testing facility, Lab. report number, Date report issued	Active ingredient (% a.i.)	#####-##	DER summary and conclusions	Core _____ Document # (assigned to DER after completion.)

2. Materials and Methods

This section of the DER should contain a description of study conduct and should include the following:

a. **Test Animals** (see Sections III.B. and C. above): species, strain, supplier, age and body weight range per sex at start of test substance administration for first generation parental animals and at Week 0 for second generation (F1) study animals, animal husbandry practices that differ from GLP requirements.

b. **Test Compound** (see Section III. A. above): purity (% active ingredient), density (if provided), physical description of the material, lot number or batch number, supplier, and date of receipt. If a list of contaminants is available, it should be included in a Confidential Business Information appendix to the DER.

c. **Test diet (or dose solution) preparation and administration** (see Sections III. D. and E.): description of any vehicle used, including information on purity, density, description, lot number, supplier, and date of receipt; information on frequency of preparation, storage conditions for test substance, vehicle and formulations, volume mixed, the use of corrections for test substance purity, method for calculating amount of test compound to be used (based on body weight for which premating, gestation or lactation day or days), results of analysis for stability, concentration, and homogeneity if available, the basis for selection of dose levels (reference to a range-finding study if available or a brief description of a range-finding study).

d. **Study Design** (see Section III. E. above): dosing schedule, dose levels, and group assignment of animals. A description of the selection of pups for the second generation should be provided.

Test Group[a]	Dose Level (ppm)[b]	Number Assigned			
		Males		Females	
		P	F1	P	F1
Control					
Low dose					
Mid dose					
High dose					

a If more than one control or middle dose group, indicate separation (e.g., low-mid dose group, high-mid dose group, etc.).

b Mg/kg/day if test compound is administered by gavage; mg/l if administered by inhalation.

e. **Mating Procedures** (see Section III. F. above): description of the type of mating (male:female ratio), criteria used to determine success of mating (e.g., presence of vaginal plug or presence of sperm in vaginal smears), length of mating period, method of exchanging mating pairs if there is no evidence of copulation.

f. **Observations** (see Sections IV. A. and B. above): descriptions of methods, schedules, and procedures used to generate all parental and filial end points, including, as appropriate, but not limited to: clinical and litter observations (including malformations), body weight, food consumption, mating and gestation length data, litter size and survival information, sacrifice, gross pathology, organ weights, tissue collection, and histopathology; if applicable, the use of 10% ammonium sulfide to detect implantation sites in supposedly pregnant animals should be mentioned.

g. **Historical Control Data** (see Section IV. B. 11.): if provided for comparison with concurrent controls and treated groups, description of the number of studies used to compile the historical data base and a list of specific end points included.

h. **Statistical analysis** (see Section III. H.): description of the statistical tests that were applied to each type of data analyzed, state level of significance used and whether tests were interpreted with one-way or two-way variance, the calculations used for reported reproductive indices, a discussion of any exclusions, a mention of the use of the litter as the basic unit of analysis, and comments on acceptability of the procedures.

3. Reported Results

This section of the DER should be divided into separate discussions of the results from the observation of (1) **parental end points** (see Section IV. A.) and (2) **offspring end points** (see Section IV. B.). For each data type, the DER should clearly and concisely discuss all meaningful findings, statistical significances, deviations from guideline, and interpretations of the data as they relate to test substance toxicity.

Numerical data should be summarized in the DER, particularly when it is important in defining a NOEL, determining the adequacy of the doses tested, or supporting any conclusions about the reproductive toxicity potential of the test material. Both parental and filial endpoints should be clearly identified by generation, study phase, and sex. It is important to include the number of animals, pups, and litters examined in each group along with any descriptive statistics (i.e., mean, standard deviation, standard error, ratio, index) for each set of significant data, since the group size may change according to the end point considered and since statistical significance is not always consistent with biological significance. Statistically significant differences should be footnoted with the appropriate *p* value and name of the statistical test applied. Data tables should indicate the source (laboratory study number and report page number) of the data which were extracted. Calculations that were performed by the Reviewer should be so identified, and methods should be clearly described.

4. Discussion

This section of the DER should contain discussions of pertinent study findings, a description of the study NOEL(s) and LOEL(s) as well as the specific parental and filial findings upon which they were based, differences in data interpretations made by the investigators and the

reviewer, a list of all study deficiencies and problems, requests for additional information that may be provided by the study sponsor or laboratory, and a Core classification (see Section V. 5., below) with the basis for any Core "Supplementary" or "Invalid" study classification.

5. Core Classification System

The Core classification system is a procedure by which studies are graded with regard to the adequacy of study design, conduct, and reporting, with no regard to scientific outcome (Engler and Quest, 1988). The four categories are defined as follows:

Core Guideline: the study totally conforms with Subpart F Guidelines (§83-4).

Core minimum: the study is sufficient to fulfill the intent of the Guidelines.

Core supplementary: the study does not meet minimum criteria even though it may contain scientifically useful information.

Invalid: the study is seriously and usually irreparably flawed with respect to study design and/or scientific content and cannot be used for regulatory purposes.

Studies classified as Guideline or Minimum are considered to be acceptable for regulatory purposes, whereas those given classifications of Supplementary or Invalid are generally not. Core Supplementary studies may be conditionally upgraded following submission, review, and approval of additional information which was not included with the original study submission. Invalid studies can seldom be upgraded.

B. Outline for Peer Review Committee Presentations

The reviewer should refer to the Guidelines for Reproductive Toxicity Risk Assessment (USEPA, 1993a) when preparing a Peer Review Document for Reproductive Toxicity. The Guidelines indicate that an assessment of a pesticide or any other chemical includes the following three parts:

1. **Hazard identification/dose-response** involves the evaluation of all available experimental animal and human data and the associated doses, routes and durations of exposure to determine if an agent causes reproductive toxicity in that species and under what exposure conditions.

2. **Exposure assessment** in which the exposed population and conditions of exposure are described.

3. **Risk characterization** in which parts 1. and 2., are combined to estimate some measure of the risk of reproductive toxicity.

The Peer Review Document should fulfill part 1. The document should be formatted in a manner similar to that of the material submitted to the Developmental Toxicity or Carcinogenicity Peer Review committee. The document should follow the outline given below.

Suggested Outline for Reproductive
Toxicity Peer Review Committee Presentations

I. Introduction

(Includes a brief description of the uses for the chemical and its chemical names, synonyms and structure.)

II. Qualitative Assessment of Relevant Data

A. Rat Study #1

1. Description of maternal toxicity (including data to show dose-response and extent of effects). This should provide the Committee with sufficient information to arrive at conclusions regarding the appropriateness of dose selection.

2. Description of reproductive toxicity (including data to show dose-response and type of reproductive effects). This should provide the Committee with sufficient information to arrive at conclusions.

3. Summary of deficiencies and limitations of the study.

B. Rat Study #n (if available)

1. (Same as A. 1., above)
2. (Same as A. 2., above)
3. (Same as A. 3., above)

C. Other Species Study #n (same as A. and B., above) (if available)

III. Other Data

A. Developmental Studies

B. Subchronic and Chronic Toxicity Data

C. Mutagenicity Studies

D. Metabolism/pharmacokinetics/physico-chemical Data

E. Structure-Activity Relationships

IV. Strength of the Evidence

A. Strength of the Evidence

1. the quality of the data,
2. the resolving power of the studies,
3. the number and types of endpoints examined,
4. the relevance of route and timing of exposure,

5. the appropriateness of dose selection,
6. the reproducibility of the effects,
7. the number of species examined,
8. pharmacokinetic data,
9. structure-activity relationships, and
10. other factors

B. Questions to the Committee

A minimum of two appendices should follow the document outlined above. They should contain the DERs for the reproductive toxicity and developmental toxicity studies (Appendix 1) and the Toxicology "One-Liners" (Appendix 2). Additional appendices may be also needed, e.g., historical control data.

VII. BIBLIOGRAPHY

Adams, J., Buelke-Sam J., Kimmel C.A., Nelson C. J., Reiter L.W., Sobotka T.J., Tilson H.A., and Nelson B.K. (1985). Collaborative behavioral teralology study; Protocol design and testing procedure. Neurobehavioral Toxicology and Teratology, 7:579-586.

Adler, N.T., Resko, F.A. and Goy, R.W., (1970). The effect of copulatory behavior on hormonal change in the female rat prior to implantation. Physiol. Behavior., 5:1003-1007.

Amann, R.P., (1981). A critical review of methods for evaluation of spermatogenesis from seminal characteristics. J. Androl. 2:37-58.

Amann, R.P., (1986). Detection of alterations in testicular and epididymal function in laboratory animals. Environ. Health Perspect. 70: 149-158 (as cited in NRC Biological Markers, 1989).

Anver, M.R. and Cohen, B.J.,(1979). Lesions associated with aging. In "The Laboratory Rat", eds., HJ. Baker, JR. Lindsey, and SH. Weisbroth, Vol. 1, Academic Press.

Baetjer, A.M.,(1968). Role of environmental temperature and humidity In susceptibility to disease. Arch. Environ. Health, 16:565-570.

Baker, H.J., Lindsey, J.R. and Weisbroth, S.H., (1979). Housing to control research variables. In "The Laboratory Rat", eds., Baker H.J., Lindsey J.R., and Weisbroth S.H., Vol. 1, Academic Press.

Ball et al., (1947).

Bellhorn, R.W., (1980). Lighting in the animal environment. Lab. Animal Sci., 30(2):440-450.

Berg, B.N., (1965). Dietary restriction and reproduction in the rat. J. Nutrition, 87:344-348.

Blazak, W.F., Ernst, T.L., and Stewart, B.E., (1985). Potential indicators of reproductive toxicity: testicular sperm production and epididymal sperm number, transit, time, and motility in Fischer 344 rats. Fundam. Appl. Tox., 5:1097-1103.

Brunner, R.L., Voorhees, C.V., Kinney, L., and Butcher, R.E., (1979). Aspartame: assessment of developmental psychotoxicity of a new artificial sweetner. Neurobehav. Tox., 1:79-86.

Butcher, R.E., Voorhees, C.V., and Wooten, V., (1984). Behavioral and physical development of rats chronically exposed to caffeinated fluids. Fundam. Appl. Tox., 4:1-13.

Cambon et al., (1979).

Chapin, R.E., Gulati, D.K., Fail, P.A., Hope, E., Russell, S.R., Heindel, J.J., George, J.D., Grizzle, T.B., and Teague, J.L. (1993a). The effects of feed restriction on reproductive function in Swiss CD-1 mice. Fundam. Appl. Toxicol. 20: 15-22.

Chapin, R.E., Gulati, D.K., Barnes, L.H. and Teague, J.L. (1993). The effects of feed restriction on reproductive function in Sprague-Dawley rats. Fundam. Appl. Toxicol. 20: 23-29.

Cheeke, P.R. and Patton, N.M. (1982). In "Rabbit Production", Inter-state Printel-s and Publishers, Inc.

Chernoff, N. and Kavlock, R.J. (1982) An in vivo teratology screen utilizing pregnant mice. J. Toxicol. Environ. Health 10: 541-550.

Chester, R.V. and Zucker, I. (1970). Influence of male copulatory behavior on sperm transport, pregnancy, and pseudopregnancy in female rats. Physiol. Behavior., 5:35-43.

Christian, M.S., (1986). A critical review of multigeneration studies. J. Amer. Coll. Tox., 5:161-180.

Clegg, D.J. (1979). Animal Reproduction and Carcinogenicity Studies in Relation to Human Safety Evaluation in Toxicology and Occupational Medicine, eds., Diechman, W.B. Elsevier, North-Holland Biomedical Press, New York. pp. 44-59.

Clemens, G., Fort, F., Olberholtzer, K., Lamb, I. (1992). Midwest Teratology Association (MTA) Historical Control Database Survey (HCDS), Phase 1 (P1): Maternal Reproductive (Segment II) Indices. Poster presented at Teratology Society Annual Meeting.

Collins, T.F.X., (1978a). Reproduction and developmental toxicity guidelines: Review of deliberations by the National Toxicology Advisory Committee's reproductive panel. J. Environ. Pathol. Toxicol., 2:141-47.

Collins, T.F.X., (1978b). Multigeneration reproduction studies. In "Handbook of Teratology", Vol. 4, eds., Wilson J.G. and Fraser F.C.,Plenum Press, New York.

Darney, (1990).

Davison, J.M. and Hytten, F.E. (1974). Glomerular filtration rate during and after pregnancy. J. Obstet. Gynaecol. Gr. Commonw. 82:588 (as cited in Krauer, 1987).

Degraeve, N. (1981). Carcinogenic, teratogenic, and mutagenic effects of cadmium. Mutation Res. 86:115-135.

de Kretser, D.M. and Kerr, J.B. (1988). The cytology of the testis. In: "The Physiology of Reproduction", eds., Knobil, E., and Neil, J., et al., Raven Press, Ltd., New York, pp. 837-932.

Dixon, R.L. and Hall, J.L., (1982). Reproductive toxicology. In: "Principles and Methods of Toxicology", ed., Hayes. A.W., Raven Press, New York.

Engler,R. and Quest J.A. (1988). Review of the Office of Pesticide Programs current toxiclogy guidelines for the testing of pesticides. Tox. and Indus. Health., 4(3)i 323-329.

Food and Agriculture Organization of the United Nations (FAO), Working Party on Pesticide Residues. (1990) Pesticide Residues in Food., World Health Organization., Geneva.

Feussner, E.L., Lightkep, G.E., Hennesy, R.A., Hoberman, A.M. and Christian, M.S., (1992). A decade of rabbit fertility data: study of historical control animals. Teratology 46: 349-365.

Foote, (1986). Use of quantitative testicular histology to assess the effect of dibromochloropropane (DBCP) on reproduction in rabbits. Fund App Tox 6:638-647

Fitzhugh, O.G., (1968). Reproductive tests. In "Modern Trends in

Toxicology", eds., Boyland, E. and Goulding, R., Vol. 1, Butterworth, London.

Fox, J.G. ed. (1988). Biology and Diseases of the Ferret, Lea and Febiger, Philadelphia, PA, p. 181.

Frisch, R.E., 1978. Population, food intake, and fertility. Science. 199:22-30.

Gabrielson and Larson

Gad, S., and C. S. Weil. (1986). Statistics and Experimental Design for Toxicologists. The Telford Press, Caldwell, New Jersey.

Giovanni, E., et al., (1986). Effect of method of administration on the teratogenicity of dinoseb in the rat. Arch. Environ. Contam. Toxicol. 15:377-384.

Goldenthal, E.I., (1966). Guidelines for Safety Evaluation of Drugs for Human Use. Drug Review Branch, Division of Toxicological Evaluation, Bureau of Science, Food and Drug Administration, Washington, DC.

Gray, L.E., Ostby, L., Linder, R., Goldman, J., Rehnberg, G. and Cooper, R., (1990). Carbendizim-induced alterations of reproductive development and function in the rat and hamster. Fund Appl Tox 15: 281-297.

Gray, L.E., Ostby,J., Sigmon,R., Ferrell, J., Rehnberg, G., Linder, R., Cooper, R., Goldman, J. and Laskey, J. (1988). The development of a protocol to assess reproductive effects of toxicants in the rat. Reprod. Toxicol. 2: 281-287.

Green, S., Auletta, A., Fabricant, J., Kapp, R., Manandhar, M., Sheu, C., Springer, J.A. and Whitfield, B. (1985). Current status of assays in genetic toxicology, the dominant lethal test: a report of the U.S. Environmental Protection Agency Gene-Tox Program. Mutation Research 154: 49-67.

Green, S., Lavappa, K.S., Manandhar, M., Sheu, C., Whorton, E. and Springer, J.A. (1987). A guide for mutagenicity testing using the dominant lethal assay. Mutation Research 189: 167-174.

Gulati, D.K., Hope, E., Teague, J. and Chapin, R.E., (1991). Reproductive toxicity assessment by continuous breeding in Sprague-Dawley rats: A comparison of two study designs. Fund. Appl. Toxicol. 17:270-279.

Hafez, E.S.E., (1970). Reproduction and breeding techniques for laboratory animals, Lea and Febiger, Philadephia, p.3.

Hard, E., and Larsson, K. (1975). Development of air righting in rats. Brain Behavior and Evolution, 11:53-59.

Harkness, J.E. and Wagner, J.E., (1989). In "Biology and medicine of rabbits and rodents", Lea and Febiger, Philadelphia.

Harris, M.W., Chapin, R.E., Lockhart, A.C. and Jokinen, M.P. (1992). Assessment of a short-term reproductive and developmental screen. Fundam. Appl. Toxicol. 19: 186-196.

Haseman and Soares, (1976). The distribution of fetal death in control mice and its implications on statistical tests for dominant lethal effects. Mut Res 41:277-288.

Hastings, J.W. and Manaker, A.H., (1976). Physiological and biochemical aspects of circadian rhythms. Fed. Proc. Am. Soc. Exp. Biol. 35:2325-2357.

Hayes, A.W. ed. (1982). Principles and Methods of Toxicology. Raven Press, New York.

Hellwig, J. (1989). Report on the prenatal toxicity studies with reg. no. 83 258 (vinclozolin) in rats after oral adminstration (gavage)--consisting of report nos. 89/0090, 89/0091, 89/0092, 89/0093. BASF Corp. Chemicals Div., Ag. Chem., 100 Cherry Hill Road, Parsippany, NJ 07054 (MRID No. 411322-01).

Holmes, D.D., (1985). The mongolian gerbil in biomedical research. Lab. Animal Sci. 14:23-38.

IARC, (1990).

International Conference on Male-Mediated Developmental Toxicity: Father's Exposure and Their Children's Health, (1992). University of Pittsburgh Medical Center, Graduate School of Public Health, September 16-19.

Jacoby, R.O. and Fox, J.G., (1984). Biology and diseases of mice. In "Laboratory Animal Medicine", eds., Fox, J.G., Cohen, B.J. and Loew, F.M. Academic Press, Orlando, FL.

Jelovsek, F.R., Mattison, D.R., and Young, J.F., (1990). Eliciting principles of hazard indentification from experts. Teratology 42:521-533.

Johnson, E.M., (1986). The Scientific Basis for Multigeneration Safety Evaluations J. of the American College of Toxicology 5(4):31.

Kavlock, R.J.,(1991). Symposium on application of pharmacokinetics in developmental toxicity risk assessments. Fund. Appl. Toxicol. 16(4):213-232.

Kavlock (?)

Khera, K., Grice, H.C. and Clegg, D.,(1989). Interpretation and extrapolation of reproductive data to establish human safety standards, Current Issues in Toxicology, Spring-Verlag, New York, p.29.

Kimmel, C.A. and Price, C.J., (1990). Developmental toxicity studies. In: "Handbook of In Vivo Toxicity Testing", eds., Arnold, D.L., Grice, H.C., Krewski, D.R., Academic Press, San Diego, CA pp. 271-301.

Kohn, D.F. and Barthuld, S.W., (1984). Biology and diseases of rats. In "Laboratory Animal Medicine", eds., Fox,J.G., Cohen, B.J. and Loew, F.M., Academic Press, Orlando, FL.

Korenbrot, C.C., Huhtaniemi, I.T., and Weiner, R.W., (1977). Preputial separation as an external sign of pubertal development in the male rat. Biology of Reproduction, 17:298-303.

Krauer, B., (1987). Physiological changes and drug disposition during pregnancy. In: "Pharmacokinetics in Teratogenesis", Volume I, Interspecies Comparison and Maternal/Embryonic-Fetal Drug Transfer, eds., Nau, H. and Scott, Jr., W.J., CRC Press, Inc., Boca Raton, Florida.

Lamb, J.C. and Chapin, R.E., (1985). Experimental models of male reproductive toxicology. In: Endocrine Toxicology, eds., Thomas, J.A., Korach, K.S., McLachlan, J.A., Raven Press, New York. pp. 85-115.

Lee, I.P., (1983). Adaptive biochemical repair response toward germ cell DNA damage. AM. J. Ind. Med. 4:861-867 (as cited in Mattison and Thomford, 1987).

Lee, I.P. and Dixon, R.L., (1978). Factors influencing reproductive and genetic toxic effects on male gonads. Envir. Hlth. Perspect., 24:117-127 (as cited in

Mattison and Thomford, 1987).

Mattison, D.R., (1983). The mechanisms of action of reproductive toxins. In: "Reproductive Toxicology", ed., R.L. Dixon, Raven Press, New York, pp. 109-130 (as cited in Mattison and Thomford, 1987).

Mattison, D.R., (1984). How everyday drugs can affect reproduction. Contemp. Obstet. Gynec., 24:92-108 (as cited in Mattison and Thomford, 1987).

Mattison, D.R., Blann, E., and Malek, A., (1991). Physiological alterations during pregnancy:impact on toxicokinetics. Fund. Appl. Toxicol. 16(4):215-218.

Mattison, D.R. and Thomford, P.J., (1987). Selection of animals for reproductive toxicology studies: An evaluation of selected assumptions in reproductive toxicity testing and risk assessment. In: "Human Risk Assessment--The Role of Animal Selection and Extrapolation", eds., M.V. Roloff, A.G.E. Wilson and W.E. Ribelin et al., Taylor & Francis, New York, pp. 213.

Merry, B.J. and Holehan, A.M., (1979). Onset of puberty and duration of fertility in rats fed a restricted diet. J. Reprod. Fertility, 57:253-259.

Morrissey, R.E., Lamb, J.C., Morris, R.W., Chapin, R.E., Gulati, D.K., and Heindel, J.J., (1989). Results and evaluations of 48 continuous breeding reproduction studies conducted in mice. Fund. Appl. Toxicol. 13:747-777.

Mosher, W.D., Pratt, W.F., (1985). Fecundity and Infertility in the United States, 1965-1982. Demography 22(3):415-430.

Mosteller, F., and Youtz, C., (1961). Tables of the Freeman-Tukey transformation for the binomial and Poisson distributions. Biometrics. 48:433-440.

Mulder, J.B., (1971). Animal behavior and electromagnetic energy waves. Lab. Animal Sci. 21:389-393.

National Research Council (NRC), (1977). Reproduction and Teratogenicity Tests. In "Principles and Procedures for Evaluating the Toxicity of Household Substances", National Academy of Sciences.

National Research Council (NRC), (1978). Nutrient requirements of laboratory animals, 3rd revised ed., Vol. 10, National Academy of Sciences.

National Research Council (NRC), (1989). Biologic markers in reproductive toxicology:Subcommittee on reproductive and neurodevelopmental toxicology committee on biological markers. National Academy Press, Washington D.C.

National Research Council (NRC), (1992). Pharmacokinetics: Defining dosimetry for risk assessment, National Academy of Sciences workshop, Washington, D.C. March 4-5, 1992.

Nau, H., (1991). Pharmacokinetic considerations in the design and interpretation of developmental toxicity studies. Fund. Appl. Toxicol. 16(4):219-221.

Nau, H. and Scott, Jr., W.J., (1987). Pharmacokinetics in Teratogenesis, Volume I, Interspecies Comparison and Maternal/Embryonic-Fetal Drug Transfer. CRC Press, Inc., Boca Raton, Florida.

Nelson, J.F. and Felicio, L.S., (1984). Dietary modulations of estrous cyclicity in singly and multiply housed C57BL/6J mice. Lab. Animal Science, 34(2):173-176.

Newman, L.M., Johnson, E.M., and Cadogan, A.S.A., (1983). Oxygen consumption and survival prediction in neonatal rats exposed to prenatal hypervitaminosis A. Teratology. 28:219-227.

Newman, L.M., Johnson, E.M., and Roth, J.M., (1984). Lung volume and compliance in neonatal rats. Lab Animal Science 34(4):371-375.

Organization for Economic and Co-operative Development (OECD), (1983). Guideline for testing of chemicals. Guideline 416. "Two-generation reproduction toxicity study".

Organization for Economic and Cooperative Development (OECD), (1992). Screening Information Data Set (SIDS), Guideline for testing of chemicals. Guideline 421. "Reproduction/Developmental Toxicity Screening Test". Final Draft Document, TPG/92.199

Pakes, S.P., Lu, Y.S. and Meunier, P.C., (1984). Factors that complicate animal research. In "Laboratory Animal Medicine", eds., Fox, J.G., Cohen, B.J. and Loew, F.M., Academic Press, Orlando,FL.

Palmer, A.K., (1986). A simpler multigeneration study. Poster presented at the 6th International Congress of Pesticide Chemistry, Ottowa.

Palmer, A.K., (1981). Regulatory Requirements for Reproductive Toxicology: Theory and Practice. In: "Developmental Toxicology", eds., Kimmel, C.A.and Buelke-Sam, J., Raven Press, New York. pp.259-287.

Pederson, R.A. and Mangia, F., (1978). Ultraviolet light induced unscheduled DNA synthesis by resting and growing oocytes. Mutation Res. 49:425-429.

President's Science Advisory Committee, (1963).

Ribeiro, P.L., and Faustman, E.M., (1989). Pharmacokinetic considerations in developmental toxicology. Presented at ILSI Workshop on Risk Assessment Issues in Reproductive and Developmental Toxicology, September 18-19, 1989, Berkeley, California.

Reference for Fischer 344 rats

Robaire, B., (1992). Post-testicular effects of toxicants. Presentation at International Conference on Male-Mediated Developmental Toxicity: Father's Exposure and Their Children's Health, 1992. University of Pittsburgh Medical Center, Graduate School of Public Health, September 16-19.

Robaire, B., Smith, S., Hales, B.F., (1984). Suppression of spermatogenesis by testosterone in adult male rats: effect on fertility, pregnancy outcome and progency. Biol. Reprod. 31:221-230.

Robertson, R.T., Majka, J.A., Peter, C.P., and Bokelman, D.L., (1980). Effects of prenatal exposure to chlorpromazine on postnatal development and behavior of rats. Toxicology and Applied Pharmacology, 53:541-549.

Robinson, D.G., (1979). Endocrinological research. The Gerbil Digest, 6:1-4

Ruddon, R.W., (1990). Chemical Teratogenesis. In: "Principles of Drug Action, The Basis of Pharmacology", eds., Pratt, W.B., and Taylor, P., Churchill Livingstone, New York.

Russell, L.D., Ettlin, R., Sinha Hikim, A.P., and Clegg, E.D., (1990). Histological and Histopathological Evaluation of the Testis. Cache River Press, Clearwater, FL.

Schwartz, N.B., Dierschke, D.J., McCormack, C.E. and Waltz, P.W., (1977). Feedback regulations of reproductive cycles in rats,

Schwetz, B.A., Rao, K.S., Park, C.N., (1980). Insensitivity of tests for

reproductive problems. J. Environ. Pathol. Toxicol. 3:81-98.

Shapiro, (1980).

Silbergeld, E.K., (1992). Neurobehavioral outcomes. Presentation at International Conference on Male-Mediated Developmental Toxicity: Father's Exposure and Their Children's Health, University of Pittsburgh Medical Center, Graduate School of Public Health, September 16-19.

Stevens, K.R., and Gallo, M.A., (1989). Practical considerations in the conduct of chronic toxicity studies. In: "Principles and Methods of Toxicology", ed., Hayes, A.W., Raven Press, New York. pp. 237-250.

Taber, C.W., (1970). Taber's Cyclopedic Medical Dictionary. 11th Edition, F.A. Davis Company, Philadelphia.

Tardiff, R.G. and Rodricks, J.V.,(1987). Toxic substances and human risk: principles of data interpretation, Plenum Press, New York.

Thiessen, D.D. and Yahr, P., (1977). The gerbil behavioral Investigations: mechanisms of teratorality and olfactory communication., U. of Texas Press, Austin, TX.

Trasler, J.M., (1992). Paternal exposure to the demethylating drug 5-azacytidine alters pregnancy outcome and the male reproduction system. Presentation at International Conference on Male-Mediated Developmental Toxicity: Father's Exposure and Their Children's Health, 1992. University of Pittsburgh Medical Center, Graduate School of Public Health, September 16-19.

U.S. Congress, Office of Technical Assessment (OTA)., (1988). Reproductive Health Hazards in the Workplace, Chapter 4. Evidence of workplace hazards to reproductive function. J.B. Lippincott Company, Philadelphia.

United States Environmental Protection Agency (USEPA) (1982). Subpart 83-4:Reproductive and Fertility Effects. In: Pesticide Assessment Guidelines, Subdivision F, Hazard Evaluation: Human and Domestic Animals, Office of Pesticides and Toxic Substances, Washington, D.C., EPA-540/a-82-025.

United States Environmental Protection Agency (USEPA), (1985a). Toxic Substances Control Act, Part 798-Health Effects Testing Guidelines. Subpart E--Specific Organ/Tissue Toxicity, 798.4700, Reproduction and fertility effects. Fed.Reg. 50(188):39432-39433.

United States Environmental Protection Agency (USEPA), (1985b). Hazard evaluation division standard evaluation procedure. Teratology studies. Office of Pesticide Programs, Washington, DC.

United States Environmental Protection Agency (USEPA) (1986a). Guidelines for carcinogen risk assessment, Federal Register 51(185):33992-34003.

United States Environmental Protection Agency (USEPA), (1986b). Guidelines for mutagenicity risk assessment. Federal Register 51(185):34006-34012.

United States Environmental Protection Agency (USEPA), (1986c). Guidelines for estimating exposures. Federal Register 51(185):34042-34054.

United States Environmental Protection Agency (USEPA), (1986d). PR Notice 86-5, Standard format for data submitted underthe Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and certain provisions of the Federal Food, Drug, and Cosmetic Act (FFDCA), Issued July 29, 1986.

United States Environmental Protection Agency (USEPA), (1987). Workshop on One- vs. Two-Generation Reproductive Effects Studies, Sponsored by the USEPA Risk Assessment Forum, Washington, DC.

United States Environmental Protection Agency (USEPA), (1988a). Proposed guidelines for assessing male reproductive risk. Federal Register 53:24850-24869.

United States Environmental Protection Agency (USEPA), (1988b). Proposed guidelines for assessing female reproductive risk. Federal Register 53:24834-24847.

United States Environmental Protection Agency (USEPA), (1988c). Health Effects Division Standard Evaluation Procedure. Inhalation Toxicity Testing. Office of Pesticide Programs, Washington, DC. EPA-540/09-88-101.

United States Environmental Protection Agency (USEPA), Office of Drinking Water Health Advisories, (1989a). Drinking Water Health Advisory: Pesticides. Lewis Publishers, Michigan.

United States Environmental Protection Agency (USEPA), (1989b). FIFRA Accelerated Reregistration Phase 3 Technical Guidance, Office of Pesticides and Toxic Substances, Washington, DC. EPA No. 540/09-90-078.

United States Environmental Protection Agency (USEPA), (1990). Health Effects Division Peer Review of Procymidone. Dated October 31, 1990. Office of Pesticide Programs.

United States Enviromental Protection Agency (USEPA), (1991a). Pesticide Assessment Guidelines, Subdivision F, Hazard Evaluation: Human and Domestic Animals, Addendum 10, Neurotoxicity, Series 81, 82, and 83, March 1991, EPA 540/09-91-123, PB 91-154617.

United States Environmental Protection Agency (USEPA), (1991b). Guidelines for developmental toxicity risk assessment. Fed.Reg. 56(234):63798-63826.

United States Environmental Protection Agency (USEPA), (1992). Memorandum of September 23, 1992, "Atrazine Two-Generation Study", HED Ad hoc Committee for Atrazine Reproductive Issue, for a discussion of the setting of a NOEL in a reproduction study with varying mean litter sizes.

United States Environmental Protection Agency (USEPA), (1992). Code of Federal Regulations, Title 40 (40CFR)--Protection of Environment, Subchapter E--Pesticide Programs, Part 158, Data requirements for registration, pp. 88-142.

United States Environmental Protection Agency (USEPA), (1993a). Draft guidelines for reproductive toxicity risk assessment.

United States Environmental Protection Agency (USEPA), (1993b). Health Effects Division Draft Standard Evaluation Procedure. Developmental Toxicity Studies. Office of Pesticide Programs, Washington, DC.

United States Food and Drug Administration (USFDA), (1959). Appraisal of the safety of chemicals in foods, drugs and cosmetics. Published by The Association of Food and Drug Officials of the United States, Third Printing 1975.

Vessel et al., (1976).

Waller, D.P., Killinger, J.M., Zaneveld, L.J.D., (1985). Physiology and toxicology of the male reproductive tract. In: "Endocrine Toxicology, eds., Thomas, J.A., Korach, K.S., McLachlan, J.A., Raven Press, New York, NY, pp. 269-333.

Whorton, D. and Milby, T.H., (1980). Recovery of testicular function among DPCP workers. J. Occup. Med. 22:177-179.

Wiehe, (1976).

Williams, R.H. ed., (1974). Textbook of Endocrinology. W.B. Saunders Company, Phildelphia.

Working, P.K. (1988). Male reproductive toxicity: Comparison of the human to animal models. Environ. Health Perspect. 77:37-44.

World Wildlife Fund (WWF), (1992). Symposium on Endocrine Disruptors: The Wildlife/Human Connection. Washington, D.C., March 18-19

Wyrobek, A.J., Gordon, L.A., Burkhart, J.G., Francis, M.W., Kapp, R.W., Letz, G., Malling, H.V., Topham, J.C. and Whorton, D.M., (1983a). An evaluation of the mouse sperm morphology test and other sperm tests in nonhuman mammals. Mutat. Res. 115:1-72.

Wyrobek, A.J., Gordon, L.A., Burkhart, J.G., Francis, M.W., Kapp, R.W., Letz, G., Malling, H.V., Topham, J.C. and Whorton, D.M., (1983b). An evaluation of human sperm as indicators of chemically induced alterations of spermatogenic function. Mutat. Res. 115:73-148.

Yagil et al., (1976).

Yamauchi et al., (1981).

Zenick, H. and E.D. Clegg., (1989). Assessment of male reproductive toxicity: a risk assessment approach, In: "Principles and Methods of Toxicology", ed. A.W. Hayes, Raven Press, Ltd., New York, NY.